



Journal of
*Risk and Financial
Management*

Computational Finance

Edited by

Lars Stentoft

Printed Edition of the Special Issue Published in
Journal of Risk and Financial Management

Computational Finance

Computational Finance

Editor

Lars Stentoft

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin



Editor

Lars Stentoft
Department of Economics,
University of Western Ontario,
Social Science Centre Room 4071
Canada

Editorial Office

MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Journal of Risk and Financial Management* (ISSN 1911-8074) (available at: https://www.mdpi.com/journal/jrfm/special_issues/CompFin).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. <i>Journal Name</i> Year , Article Number, Page Range.

ISBN 978-3-03936-966-9 (Hbk)

ISBN 978-3-03936-967-6 (PDF)

© 2020 by the authors. Articles in this book are Open Access and distributed under the Creative Commons Attribution (CC BY) license, which allows users to download, copy and build upon published articles, as long as the author and publisher are properly credited, which ensures maximum dissemination and a wider impact of our publications.

The book as a whole is distributed by MDPI under the terms and conditions of the Creative Commons license CC BY-NC-ND.

Contents

About the Editor	vii
Preface to “Computational Finance”	ix
R. Mark Reesor and T. James Marshall Forest of Stochastic Trees: A Method for Valuing Multiple Exercise Options Reprinted from: <i>J. Risk Financial Manag.</i> 2020 , <i>13</i> , 95, doi:10.3390/jrfm13070095	1
Pascal Létourneau and Lars Stentoft Bootstrapping the Early Exercise Boundary in the Least-Squares Monte Carlo Method Reprinted from: <i>J. Risk Financial Manag.</i> 2019 , <i>12</i> , 190, doi:10.3390/jrfm12040190	33
Yuyang Cheng, Marcos Escobar-Anel and Zhenxian Gong Generalized Mean-Reverting 4/2 Factor Model Reprinted from: <i>J. Risk Financial Manag.</i> 2019 , <i>12</i> , 159, doi:10.3390/jrfm12040159	55
Peter A. Forsyth and Kenneth R. Vetzal Defined Contribution Pension Plans: Who Has Seen the Risk? Reprinted from: <i>J. Risk Financial Manag.</i> 2019 , <i>12</i> , 70, doi:10.3390/jrfm12020070	77
Purba Mukerji, Christine Chung, Timothy Walsh and Bo Xiong The Impact of Algorithmic Trading in a Simulated Asset Market Reprinted from: <i>J. Risk Financial Manag.</i> 2019 , <i>12</i> , 68, doi:10.3390/jrfm12020068	105
Lars Stentoft Efficient Numerical Pricing of American Call Options Using Symmetry Arguments Reprinted from: <i>J. Risk Financial Manag.</i> 2019 , <i>12</i> , 59, doi:10.3390/jrfm12020059	117
Peter G. Dunne Positive Liquidity Spillovers from Sovereign Bond-Backed Securities Reprinted from: <i>J. Risk Financial Manag.</i> 2019 , <i>12</i> , 58, doi:10.3390/jrfm12020058	143
Vladimir Petrov, Anton Golub and Richard Olsen Instantaneous Volatility Seasonality of High-Frequency Markets in Directional-Change Intrinsic Time Reprinted from: <i>J. Risk Financial Manag.</i> 2019 , <i>12</i> , 54, doi:10.3390/jrfm12020054	169
Johannes Stübinger, Lucas Schneider Statistical Arbitrage with Mean-Reverting Overnight Price Gaps on High-Frequency Data of the S&P 500 Reprinted from: <i>J. Risk Financial Manag.</i> 2019 , <i>12</i> , 51, doi:10.3390/jrfm12020051	201
Marcel T. P. van Dijk, Cornelis S. L. de Graaf and Cornelis W. Oosterlee Between \mathbb{P} and \mathbb{Q} : The $\mathbb{P}^{\mathbb{Q}}$ Measure for Pricing in Asset Liability Management Reprinted from: <i>J. Risk Financial Manag.</i> 2018 , <i>11</i> , 67, doi:10.3390/jrfm11040067	221

About the Editor

Lars Stentoft is an Associate Professor at the Department of Economics (joint with the Department of Statistical and Actuarial Sciences), University of Western Ontario, Canada. He has been a faculty member at Western since 2014, and prior to this, he held positions at HEC Montreal (2005–2014) and Copenhagen Business School (2012–2013). He earned his PhD degree in economics from Aarhus University in 2004. Professor Stentoft is a Research Fellow at CIRANO and an International Research Fellow at CREATES. His main fields of research are concerned with financial econometrics, where he specializes on developing flexible models for financial asset returns and their use for option pricing, and with computational finance, where he works on simulation-based methods for pricing derivatives. Professor Stentoft is an Associate Editor for the *Journal of Empirical Finance*, and on the *Editorial Board of Journal of Risk and Financial Management*. His work in financial econometrics has been published in, e.g., *International Journal of Forecasting*, *Journal of Banking and Finance*, *Journal of Empirical Finance*, and *Journal of Financial Econometrics*, and his research on American option pricing has been published in, e.g., *Journal of Computational Finance*, *Management Science*, and *Review of Derivatives Research*.

Preface to “Computational Finance”

In a Special Issue of the *Journal of Risk and Financial Management*, there was a call for contributions within the broad topic of Computational Finance. The topic includes novel research on the use of computational methods and techniques for modelling financial asset prices, returns, and volatility, and in the pricing, hedging, and risk management of financial instruments. Theoretical and empirical articles on the application of novel computational techniques in estimation, simulation, optimization, and calibration with applications to asset pricing, derivative valuation, hedging, and risk management were welcomed and contributions focusing on multivariate or high-dimensional applications in today’s complex world, novel measures of financial risk, and other types of risks implied from derivative markets, and on the use of high-frequency data of all sorts, were especially encouraged.

This book contains 10 contributions appearing in this special issue. A first group of papers contain new multivariate models, see Cheng et al. (2019), and novel techniques for pricing derivatives in such flexible models, see Reesor and Marshall (2020), and examines how flexible valuation methods based on simulation and regression techniques, see, e.g., Longstaff and Schwartz (2001), can be improved upon in general, see Létourneau and Stentoft (2019), and specifically when it comes to pricing realistic and relevant long maturity real options, see Stentoft (2019). A second group of papers examines how pricing and hedging techniques can be used to assess the challenges faced by insurance companies by changing regulatory requirements, see Van Dijk et al. (2018), by pension plan participants due to the increased use of defined contribution plans, see Forsyth and Vetzal (2019), and by market participants in general as a result of the increased sovereign linkages, see Dunne (2019). Finally, the last group of papers considers in one way or another issues related to high-frequency trading in general, see Mukerji et al. (2019), to the use of statistical arbitrage strategies in particular, see Stübinger and Schneider (2019), and to the use of high-frequency data to assess risk and volatility in financial markets, see Petrov et al. (2019).

Cheng et al. (2019) introduce a novel multivariate mean-reverting stochastic volatility model obtained by combining recently developed univariate volatility models with factor models using a principal component and factor covariance decomposition. The model is referred to as a multivariate Generalized Mean-Reverting $4/2$ Factor Model, a name that comes from the superposition of a CIR term and a $3/2$ -model component. Compared to other existing multivariate stochastic volatility models the setting used reduces the dimension of the parametric space making popular estimation methods feasible. The authors find conditions for well-defined changes of measure and develop closed form expressions for two key characteristic functions, which allow for fast and efficient calculation of derivatives prices and risk measures and exposures, and they demonstrate numerically the significant impact of their model, both with respect to changes on the implied volatility surface and on two standard risk measures.

Reesor and Marshall (2020) present a new numerical technique for pricing multiple exercise options by simulation referred to as the Forest of Stochastic Trees (FOST) method. The proposed method uses stochastic trees in place of binomial trees in the Forest of Trees algorithm originally proposed to value swing options and extends this algorithm to allow for a multi-dimensional underlying process. The method can also be viewed as extending the stochastic tree method for valuing (single exercise) American-style options to multiple exercise options. The authors

demonstrate that the proposed valuation method results in low and high biased estimators for the true option value and that these estimators are consistent for the true option value. The proposed method is of particular use in cases where there is potentially a large number of underlying assets and/or where the underlying price process depends on multiple risk factors.

Létourneau and Stentoft (2019) also consider option pricing but uses instead simulation and regression based techniques, see, e.g., Longstaff and Schwartz (2001). The authors propose an algorithm that improves on the approximation of the optimal early exercise boundary for American options. The method works by exploiting and leveraging the information in multiple cross-sectional regressions by averaging the individually obtained estimates at each early exercise step in the backwards induction algorithm. With this method, less errors are accumulated, and as a result of this, the price estimate is essentially unbiased even for long maturity options. Moreover, because the method naturally disassociates the estimation of the optimal early exercise boundary from the pricing of the option, significant efficiency gains can be obtained by using less simulated paths and repetitions to estimate the optimal early exercise boundary than with the regular method.

Stentoft (2019) also considers the LSM method of Longstaff and Schwartz (2001) and uses put-call symmetry to estimate call prices by pricing these as symmetric put options instead. His results show that, for a large sample of options with characteristics of relevance in real-life applications, the symmetric method performs much better on average, is the best method for most of the options, never performs poorly and, as a result, is extremely efficient compared to the optimal, but unfeasible method that picks the method with the smallest Root Mean Squared Error. The relative improvements of using symmetric pricing increases with option maturity and with asset volatility, and using the method to price, for example, real options, many of which are call options with long maturities on volatile assets, for example energy, could improve the estimates significantly.

Using option pricing theory, Van Dijk et al. (2018) consider the challenges faced by insurance companies when it comes to valuing issued guarantees at a one year horizon as is required by recent changes in regulation. These guarantees have option like features and the standard approach is to assume that the parameters of these option pricing models are constant, i.e., the calibrated parameters from time $t = 0$ are also used to value the guarantees at $t = 1$. However, it is well-known that the parameters are not constant and may depend on the state of the market. The authors propose an improved regression model that, given a set of market variables such as the VIX index and risk-free interest rates, can be used to estimate accurately the calibrated parameters at future dates. They show that, depending on the initial state of the market, the impact on the Solvency Capital Requirement may vary between -46% and 52% .

The increased use of Defined Contribution (DC) plans implies that more pension plan participants will bear the risk that final realized portfolio values may be insufficient to fund desired retirement cash flows. Forsyth and Vetzal (2019) compare the outcomes of various asset allocation strategies for a typical DC plan investor in a synthetic market and also using bootstrap resampling of historical data. They propose a strategy based on optimal dynamic (multi-period) time consistent quadratic shortfall and demonstrate that the probability that portfolio values at retirement will be insufficient to provide adequate retirement incomes is relatively high, unless DC investors adopt optimal allocation strategies and raise typical contribution rates. The results suggest that there is a looming crisis in DC plans, which requires educating plan holders in terms of realistic expectations, required contributions, and optimal asset allocation strategies.

The above contributions rely on having access to safe or risk free assets in combination with risky assets to generate sufficient cash flows. Dunne (2019) considers the benefits and disadvantages of introducing a European Sovereign Bond-Backed Securitisation (SBBS) to address the need for a common safe asset that would break destabilising bank-sovereign linkages. The author uses a simulation approach to assess the effectiveness of hedges incurred while making markets in individual euro area sovereign bonds by taking offsetting positions in one or more of the SBBS tranches. He finds that optimal hedging with SBBS reduces risk exposures substantially in normal market conditions. In volatile conditions, hedging is not very effective but leaves dealers exposed to mostly idiosyncratic risks which disappears if dealers are diversified in providing liquidity across country-specific secondary markets and SBBS tranches.

Mukerji et al. (2019) also use simulation methods, though at a high frequency, to study the impact of algorithmic trading (AT) in a market in which human and algorithmic counterparts trade based on technical and fundamental analysis and statistical arbitrage strategies. Considering variations in the level of market uncertainty and the degree of algorithmic versus human trading, they show that liquidity increases initially as AT rises to about 10% share of the market. The authors also demonstrate that statistical arbitrage can lead to significant deviations in asset prices from fundamentals.

Stübinger and Schneider (2019) uses a mean-reverting jump–diffusion model applied to high-frequency data of the S&P 500 constituents to identify temporary market anomalies during the first minutes of a trading day. Using data from January 1998–December 2015, the authors show that a statistical arbitrage strategy to select stocks based on overnight price gaps delivers statistically and economically significant returns of 51% and an annualized Sharpe ratio of 2.38 after transaction costs which is shown to be superior to several alternative existing quantitative strategies.

Finally, using high-frequency data from several sources Petrov et al. (2019) propose a novel intraday instantaneous volatility measure which utilises sequences of drawdowns and drawups as indicators of high-frequency activity of financial markets. The authors use their measure to uncover weekly seasonal patterns in volatility for three Forex and one Bitcoin exchange rates, as well as a stock market index. The provided volatility estimation method can be used for risk-management independent of the discreteness and the type of analysed high-frequency data.

This volume includes a wide variety of theoretical and empirical contributions that address a wide range of issues and topics related to computational finance. The published papers consider asset pricing in general with applications to bond pricing, see Dunne (2019), commodity modelling, see Cheng et al. (2019), and derivatives pricing, see Létourneau and Stentoft (2019), Reesor and Marshall (2020) and Stentoft (2019); uses calibration techniques, see Van Dijk et al. (2018); and considers issues related to hedging, see Dunne (2019). The published papers develop new multivariate models, see Cheng et al. (2019), considers option pricing in this challenging setting, see Reesor and Marshall (2020), and addresses issues related to risk management, see Cheng et al. (2019), Forsyth and Vetzal (2019), and Van Dijk et al. (2018). The volume contains several papers that use simulation, see Dunne (2019), Létourneau and Stentoft (2019), Mukerji et al. (2019), and Stentoft (2019), and papers that provide new ways to model volatility, see Cheng et al. (2019), and for estimating this, see Petrov et al. (2019).

Lars Stentoft

Editor

References

- Cheng, Yuyang, Marcos Escobar-Anel, and Zhenxian Gong. 2019. Generalized Mean-Reverting 4/2 Factor Model. *Journal of Risk and Financial Management* 12: 159.
- Dunne, Peter G. 2019. Positive Liquidity Spillovers from Sovereign Bond-Backed Securities. *Journal of Risk and Financial Management* 12: 58.
- Forsyth, Peter A., and Kenneth R. Vetzal. 2019. Defined Contribution Pension Plans: Who Has Seen the Risk? *Journal of Risk and Financial Management* 12: 70.
- Létourneau, Pasvcal, and Lars Stentoft. 2019. Bootstrapping the Early Exercise Boundary in the Least-Squares Monte Carlo Method. *Journal of Risk and Financial Management* 12: 190.
- Longstaff, Francis A., and Eduardo S. Schwartz. 2001. Valuing American Options by Simulation: A Simple Least-Squares Approach. *Review of Financial Studies* 14: 113–47.
- Mukerji, Purba, Christine Chung, Timothy Walsh, and Bo Xiong. 2019. The Impact of Algorithmic Trading in a Simulated Asset Market. *Journal of Risk and Financial Management* 12: 68.
- Petrov, Vladimir, Anton Golub, and Richard Olsen. 2019. Instantaneous Volatility Seasonality of High-Frequency Markets in Directional-Change Intrinsic Time. *Journal of Risk and Financial Management* 12: 54.
- Reesor, R. Marm, and T. James Marshall. 2020. Forest of Stochastic Trees: A Method for Valuing Multiple Exercise Options. *Journal of Risk and Financial Management* 13: 95.
- Stentoft, Lars. 2019. Efficient Numerical Pricing of American Call Options Using Symmetry Arguments. *Journal of Risk and Financial Management* 12: 59.
- Stübinger, Johannes, and Lucas Schneider. 2019. Statistical Arbitrage with Mean-Reverting Overnight Price Gaps on High-Frequency Data of the S&P 500. *Journal of Risk and Financial Management* 12: 51.
- Van Dijk, Marcel T. P., Cornelis S. L. De Graaf, and Cornelis W. Oosterlee. 2018. Between P and Q : The P^Q Measure for Pricing in Asset Liability Management. *Journal of Risk and Financial Management* 11: 67.



Article

Forest of Stochastic Trees: A Method for Valuing Multiple Exercise Options

R. Mark Reesor ^{1,*} and T. James Marshall ²

¹ Department of Mathematics, Wilfrid Laurier University, Waterloo, ON N2L 3C5, Canada

² Bank of Montreal, Toronto, ON M5X 1A1, Canada; JAMES1.MARSHALL@bmo.com

* Correspondence: mreesor@wlu.ca

Received: 31 August 2019; Accepted: 3 May 2020; Published: 13 May 2020

Abstract: We present the Forest of Stochastic Trees (FOST) method for pricing multiple exercise options by simulation. The proposed method uses stochastic trees in place of binomial trees in the Forest of Trees algorithm originally proposed to value swing options, hence extending that method to allow for a multi-dimensional underlying process. The FOST can also be viewed as extending the stochastic tree method for valuing (single exercise) American-style options to multiple exercise options. The proposed valuation method results in positively- and negatively-biased estimators for the true option value. We prove the sign of the estimator bias and show that these estimators are consistent for the true option value. This method is of particular use in cases where there is potentially a large number of assets underlying the contract and/or the underlying price process depends on multiple risk factors. Numerical results are presented to illustrate the method.

Keywords: Monte Carlo; multiple exercise options; dynamic programming; stochastic optimal control

1. Introduction

The broad class of stochastic optimal control problems includes many important applications in management sciences and quantitative finance such as development of natural resources, project initiation or abandonment, maintenance scheduling, land use decisions, and valuation and hedging of complex contracts. Very few problems have closed form solutions. For example the Black–Scholes–Merton formula for pricing European-style options does not have an American-style option analog. Approaches such as binomial lattice methods, partial differential equation (PDE) methods, variational inequalities and integral equations have been adopted for pricing these types of derivatives. However, all of these methods mentioned are limited in the number of sources of uncertainty and the dimensionality of the underlying asset that can be practically incorporated, leaving Monte Carlo (MC) as a general method of solution.

Multiple exercise options (MEOs) are generalizations of American-style options as they provide the holder more than one exercise right and sometimes control over one or more other variables, such as the amount exercised. Similar to pricing American-style options, MEO valuation is a problem in stochastic optimal control. For American-style options, the solution provides both a value and optimal exercise rule—a stopping time. For MEOs the solution gives both a value and optimal exercise policy. In cases in which the holder controls only the exercise times, the exercise policy is a sequence of stopping times. For MEOs in which the holder controls the exercise times and amounts, the exercise policy is a paired sequence of stopping times and exercise amounts. The policy generalizes to other control variables. Multiple exercise option valuation algorithms are generalizations of those used for pricing American-style options. Most of the work in the literature has focused on swing options in which the holder controls the exercise times and amounts. Hence, we discuss MEO valuation in the

context of swing options and note that the FOST method and the FOST price estimators’ properties apply to general MEOs.

The Forest of (Binomial/Trinomial) Trees method of [Lari et al. \(2001\)](#) and [Jaillet et al. \(2004\)](#) extends the binomial method of [Cox et al. \(1979\)](#) for pricing American style options to value swing options. As with most tree-based methods, the Forest of Trees is not able to handle high-dimensional underlying processes. Tangential to the Forest of Trees extension, the Stochastic Tree method of [Broadie and Glasserman \(1997\)](#) extends the binomial method for pricing American style options to allow for a high-dimensional underlying asset.

In this paper, we replace the binomial and trinomial trees of [Lari et al. \(2001\)](#) and [Jaillet et al. \(2004\)](#) with the stochastic trees of [Broadie and Glasserman \(1997\)](#), hence creating the Forest of Stochastic Trees (FOST) method for valuing multiple exercise options. The FOST can be thought of as generalizations of two different methodologies; specifically, it extends

- The Forest of Trees method of [Lari et al. \(2001\)](#) and [Jaillet et al. \(2004\)](#) for valuing multiple exercise options on a single asset to allow for high-dimensional underlying assets and processes; and
- The Stochastic Tree method of [Broadie and Glasserman \(1997\)](#) for valuing high-dimensional American-style options (single exercise right) to options having multiple exercise rights and additional controls (e.g., volume control).

Visually these generalizations complete the final entry in the two-by-two table shown in Table 1.

Table 1. Table showing that the Forest of Stochastic Trees extends (i) the Forest of Trees method to a high-dimensional underlying; and (ii) the Stochastic Tree method to multiple exercise options.

	1-Dimensional Asset	High-Dimensional Asset
Single exercise American option	Binomial Tree Cox et al. (1979)	Stochastic Tree Broadie and Glasserman (1997)
Multiple exercise option	Forest of Trees Lari et al. (2001) , Jaillet et al. (2004)	Forest of Stochastic Trees Marshall and Reesor

We construct high and low biased FOST estimators that are analogous to those from the Stochastic Tree. The main theoretical results, presented in Sections 2.2 and 2.3 and proven Appendix B, are listed below.

1. The high FOST estimator has positive bias.
2. The low FOST estimator has negative bias.
3. On any given realization the high FOST estimator is at least as big as the low FOST estimator.
4. The high and low FOST estimators are asymptotically unbiased.

The remainder of this section provides a literature review. Section 2 describes the Forest of Stochastic Trees estimators in detail; specifies results giving estimator properties (e.g., biasedness and convergence); provides numerical results showing the method is effective at pricing; and supplies discussion on improving computational performance through parallel processing. Section 3 summarizes and concludes the paper. A list of the notation used is given in Appendix A and proofs of the paper’s main theoretical contributions on estimator properties are given in Appendix B.

Literature Review

Multiple exercise options arise in many different areas and the structure of these contracts is typically tailored to particular clients/needs, in contrast to standardized derivatives such as interest rate swaps and exchange-traded commodity futures. A non-exhaustive list of examples of MEOs include (i) tolling agreements used in the steel [Kim et al. \(2019\)](#) and electricity [Deng and Oren \(2006\)](#) sectors; (ii) chooser flexible caps which are exotic interest rate derivatives [Meinshausen and Hambly \(2004\)](#); (iii)

valuation and control of energy production and storage facilities [Chen and Forsyth \(2007\)](#); [Ludkovski and Carmona \(2010\)](#); [Thompson et al. \(2009\)](#); and (iv) swing options [Calvo-Garrido et al. \(2017\)](#); [Jaillet et al. \(2004\)](#); [Lari et al. \(2001\)](#); [Wilhelm and Winter \(2008\)](#).

Valuation methods for MEOs are extensions of those used for American-style options. There are continuous-time solutions to both the American-style and multiple exercise option valuation problems; these are computed by solving a system of Hamilton–Jacobi–Bellman quasi-variational inequalities [Korn et al. \(2005\)](#). These methods give more accurate and stable price and sensitivity estimates than those computed using simpler tools (e.g., trees). However, these methods are quite complex mathematically and break down in higher dimensions.

In this article, we focus on the mathematically simpler time-discretized version of the valuation problem. Discrete-time tree-based methods for valuing American-style options [Cox et al. \(1979\)](#) have been extended to MEOs via the Forest of Trees [Jaillet et al. \(2004\)](#); [Lari et al. \(2001\)](#). Techniques for pricing American-style options using solutions of PDEs have been modified to MEOs [Calvo-Garrido et al. \(2017\)](#); [Chen and Forsyth \(2007\)](#); [Thompson et al. \(2009\)](#); [Wilhelm and Winter \(2008\)](#). These methods for MEOs inherit properties similar to the corresponding methods for single-exercise options. One crucial property is that these methods fail as the dimensionality of the problem increases.

Monte Carlo is the obvious tool to overcome the curse of dimensionality, as the rate of convergence of Monte Carlo estimators is independent of the dimension. [Tilley \(1993\)](#) was the first to show that the forward-in-time Monte Carlo approach could be used to solve the backward-in-time dynamic programming problem arising from valuation of an American-style option. Since this seminal paper, numerous other methods for the Monte Carlo valuation of American style options have appeared. These include methods that attempt to parameterize the exercise region [Barraquand and Martineau \(1995\)](#) and those that discretize the state space [Bally et al. \(2005\)](#). Methods that parameterize the early-exercise region have been extended to value multiple exercise options by parameterizing the set of exercise level curves [Ibáñez \(1996\)](#). Similarly state space aggregation methods have been used for multiple exercise option valuation [Ben Latifa et al. \(2016\)](#). These approaches, however, also suffer from the curse of dimensionality and do not easily generalize to arbitrary payoffs and underlying price processes.

Monte Carlo methods that do not break down with the dimensionality and that accommodate general payoff and price processes include those that solve the optimal stopping-time problem through estimation of the hold or continuation value. These include the stochastic tree and mesh techniques of [Broadie and Glasserman \(1997 2004\)](#) and the regression-based approach first appearing in [Carriere \(1996\)](#) and then subsequently generalized in [Longstaff and Schwartz \(2001\)](#). For each of these valuation techniques, high- and low-biased estimators are easily generated, along with a hybrid interleaving estimator that has properties of both. Duality-based methods solve the optimal control problem in the dual space by approximating an optimal martingale, typically by regression [Andersen and Broadie \(2004\)](#); [Haugh and Kogan \(2004\)](#).

Least-squares Monte Carlo has been modified for the pricing of swing options in [Barrera et al. \(2006\)](#); [Meinshausen and Hambly \(2004\)](#), respectively. Although increased dimensionality does not decrease the performance of these methods, they suffer from other drawbacks. In least-squares Monte Carlo methods one must select a set of basis functions on which to run regressions to estimate continuation values. In general only a complete (infinite) set of basis functions results in continuation value estimators that are consistent for the true option value. In practice, of course, a finite set of basis functions is used and introduces an approximation error. Continuation value estimators are consistent for the true approximation value and not the true option value [Clement et al. \(2001\)](#); [Stentoft \(2004\)](#).

Duality methods have been extended to MEOs [Bender \(2011\)](#); [Chandramouli and Haugh \(2012\)](#); [Gyurko et al. \(2015\)](#); [Meinshausen and Hambly \(2004\)](#). Duality methods rely on having a sub-optimal exercise policy that produces a low-biased estimate from which the solution to the dual problem can be approximated to yield a high-biased estimate. Typically regression-based methods are used to estimate

the sub-optimal exercise policy [Chandramouli and Haugh \(2012\)](#); [Gyurko et al. \(2015\)](#); [Meinshausen and Hambly \(2004\)](#) implying the above noted issues of least-squares Monte Carlo persist when pricing MEOs. Policy iteration methods such as [Bender \(2011\)](#), yield approximations of the time-0 value at each iteration of the dynamic program. As with the pricing of American-style options this method is advantageous because it removes the requirement to calculate nested conditional expectations prior to the time-0 value being approximated.

The stochastic mesh of [Broadie and Glasserman \(2004\)](#) has been extended to MEOs via the Forest of Stochastic Meshes (FOSM) [Marshall \(2012\)](#); [Marshall and Reesor \(2011\)](#). High and low biased FOSM estimators are derived [Marshall and Reesor \(2011\)](#) and their properties shown [Marshall \(2012\)](#), similar to the work presented here for the Forest of Stochastic Trees estimators.

2. Results

We consider the valuation of multiple exercise options as a stochastic optimal control problem with three relevant state variables—the underlying variable (S), number of exercise rights remaining (\mathcal{N}), and usage level (U) assuming some volume control. At each exercise opportunity and given (S, \mathcal{N}, U) , the current values of the state variables, the holder must choose between

- Exercising u units plus continuing with an option having $\mathcal{N} - 1$ remaining exercise rights and usage level $U + u$; and
- Continuing with an option having \mathcal{N} exercise rights and usage level U (i.e., no exercise).

Note that with volume control the payoff from exercising u units changes with u (as does the continuation value of the option). Thus, the holder chooses the value-maximizing u when deciding to exercise. Also note that with $\mathcal{N} = 1$ and u constrained to be 1, this is an American-style option.

We work with the time-discretized problem and use dynamic programming to solve for the optimal exercise policy and the corresponding optimal value. In all variables, let the subscript i denote time- t_i and let \mathcal{U}_i be the time- t_i set of admissible volume choices which includes the zero volume choice (i.e., hold). The recursive equations for the dynamic program are

$$H_i(S_i, \mathcal{N}_{i+1}, U_{i+1}) = E[B_{i+1}(S_{i+1}, \mathcal{N}_{i+1}, U_{i+1}) | \mathcal{Z}_i] \quad \text{and} \quad (1)$$

$$B_i(S_i, \mathcal{N}_i, U_i) = \max_{u \in \mathcal{U}_i} \left[h_i(S_i, \mathcal{N}_i, U_i, u) + H_i(S_i, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) \right], \quad (2)$$

with the terminal conditions

$$H_m(S_m, \mathcal{N}_m, U_m) = \tilde{\phi}(U_m) \quad \text{and} \quad (3)$$

$$B_m(S_m, \mathcal{N}_m, U_m) = \max_{u \in \mathcal{U}_m} \left[h_m(S_m, \mathcal{N}_m, U_m, u) + H_m(S_m, \mathcal{N}_m - I_{\{u \neq 0\}}, U_m + u) \right], \quad (4)$$

where $H_i(S, \mathcal{N}, U)$ and $B_i(S, \mathcal{N}, U)$ are the time- t_i , state- \mathcal{Z}_i continuation and option values, respectively, $h_i(S, \mathcal{N}, U, u)$ is the payoff from exercising u units with $h_i(S, \mathcal{N}, U, 0) = 0$, \mathcal{Z}_i is the time- t_i information set generated by the paths of (S, \mathcal{N}, U) , I is an indicator function and $\tilde{\phi}(\cdot)$ is a cumulative usage penalty term. Estimator properties and their proofs are given for this multiple exercise option setup. However, the dynamic program and estimator properties can be stated and proven for alternative specifications provided there is a finite number of exercise rights and usage levels. For example, a swing option contract may specify a certain number of *up* and *down* swing rights, \mathcal{N}_u and \mathcal{N}_d . An up swing right allows the holder to take more than the baseline amount of the underlying asset while a down swing right allows the holder to take less. Another variation is to allow for multiple rights to be exercised at each opportunity where each right corresponds to a fixed volume amount [Bender and Schoenmakers \(2006\)](#); [Meinshausen and Hambly \(2004\)](#).

2.1. Forest of Stochastic Trees

The FOST generalizes the stochastic tree method for valuing American-style options to the valuation of multiple exercise options and extends the Forest of Trees method to handle a high-dimensional underlying asset. This is done by replacing the binomial/trinomial trees with stochastic trees in the framework of [Lari et al. \(2001\)](#) and [Jaillet et al. \(2004\)](#) hence giving the FOST. The stochastic tree is constructed identically as described in [Broadie and Glasserman \(1997\)](#) and the tree is replicated multiple times, with one replication corresponding to each possible (\mathcal{N}, U) combination. This is analogous to the Forest of Trees in which the same underlying binomial/trinomial tree is replicated for each possible (\mathcal{N}, U) combination.

The dynamic program is approximately solved by replacing the continuation values in Equations (1) and (3) with stochastic tree-type estimators. As with the original stochastic tree technique, high- and low-biased option value estimators are constructed by using the analogous high- and low-biased estimators, respectively, on each stochastic tree in the forest. The recursive equations for the high estimator are

$$\hat{H}_i(\mathbf{S}_i^{\mathbf{j}}, \mathcal{N}_{i+1}, U_{i+1}) = \frac{1}{b} \sum_{k=1}^b \hat{V}_{i+1}(\mathbf{S}_{i+1}^{\mathbf{k}}, \mathcal{N}_{i+1}, U_{i+1}), \text{ and} \tag{5}$$

$$\hat{V}_i(\mathbf{S}_i^{\mathbf{j}}, \mathcal{N}_i, U_i) = \max_{u \in \mathcal{U}_i} \left[h_i(\mathbf{S}_i^{\mathbf{j}}, \mathcal{N}_i, U_i, u) + \hat{H}_i(\mathbf{S}_i^{\mathbf{j}}, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) \right], \tag{6}$$

with the terminal conditions

$$\hat{V}_m(\mathbf{S}_m^{\mathbf{j}}, \mathcal{N}_m, U_m) = \max_{u \in \mathcal{U}_m} \left[h_m(\mathbf{S}_m^{\mathbf{j}}, \mathcal{N}_m, U_m, u) + \tilde{\phi}(U_m + u) \right], \tag{7}$$

where $\hat{H}_i(\mathbf{S}, \mathcal{N}, U)$ and $\hat{V}_i(\mathbf{S}, \mathcal{N}, U)$ are the time- t_i , state- \mathcal{Z}_i continuation and option value estimators, respectively, $h_i(\mathbf{S}, \mathcal{N}, U, u)$ (with $h_i(\mathbf{S}, \mathcal{N}, U, 0) = 0$) is the time- t_i , state- \mathcal{Z}_i payoff from exercising u units, b is the branching factor, I is an indicator function and $\tilde{\phi}(U_m + u)$ is a global usage penalty term. The superscript $\mathbf{j} = \{j_0, j_1, \dots, j_i\}$ indicates the specific node within a given stochastic tree and $\mathbf{k} = \{j, k\}$.

Figure 1 is a diagram of a section of a Forest of Stochastic Trees with two volume choices, u_1 and u_2 . It illustrates the nodes in the forest which need to be considered when making an exercise decision given state (\mathcal{N}, U) . The three choices are no exercise, exercise u_1 units, and exercise u_2 units.

This argument has been suppressed to this point for convenience. We begin with the theorem regarding the bias of the high estimator.

Theorem 1. (High estimator bias) *The high estimator is biased high, i.e.,*

$$E [\hat{V}_0 (b, \mathbf{S}_0, \mathcal{N}_0, U_0)] \geq B_0 (\mathbf{S}_0, \mathcal{N}_0, U_0) \quad (13)$$

for all b .

Similarly, the result stating the bias of the low estimator follows.

Theorem 2. (Low estimator bias) *The low estimator is biased low, i.e.,*

$$E [\hat{\vartheta}_0 (b, \mathbf{S}_0, \mathcal{N}_0, U_0)] \leq B_0 (\mathbf{S}_0, \mathcal{N}_0, U_0) \quad (14)$$

for all b .

Finally, an ordering result for the high and low estimators is stated in Theorem 3.

Theorem 3. (Comparison of Estimators) *On every realization of the forest the low estimator is less than or equal to the high estimator. That is,*

$$\hat{\vartheta}_i (b, \mathbf{S}_i^j, \mathcal{N}_i, U_i) \leq \hat{V}_i (b, \mathbf{S}_i^j, \mathcal{N}_i, U_i) \quad (15)$$

with probability one for all \mathbf{j}, i .

Theorems 1–3 are proven in Appendix B.

2.3. Estimator Convergence

An advantage of the stochastic tree method over some other MC valuation methods is that its estimators are consistent to the true option value. This property also holds for the FOST estimators. In this section we state two theorems—one for the consistency of the high estimator, and the other for the consistency of the low estimator. Here convergence is in probability to the true option value and as above the argument b that appears with the estimators refers to an arbitrary branching factor size of b with convergence being shown as $b \rightarrow \infty$. Before stating the result, define $\bar{V}_0 (b, \mathbf{S}_0, \mathcal{N}_0, U_0)$ as the average of R independent replications of $\hat{V}_0 (b, \mathbf{S}_0, \mathcal{N}_0, U_0)$.

Theorem 4. (High estimator convergence) *Suppose $E [|h_i (\mathbf{S}_i, \mathcal{N}_i, U_i)|^{p'}] < \infty$, for all t_i , and some $p' > 1$. Then $\bar{V}_0 (b, \mathbf{S}_0, \mathcal{N}_0, U_0)$ converges to $B_0 (\mathbf{S}_0, \mathcal{N}_0, U_0)$ in p -norm for any $0 < p < p'$ as $b \rightarrow \infty$. This holds for an arbitrary number of repeated valuations of the forest, R . In particular $\bar{V}_0 (b, \mathbf{S}_0, \mathcal{N}_0, U_0)$ converges to $B_0 (\mathbf{S}_0, \mathcal{N}_0, U_0)$ in probability and is thus a consistent estimator of the option value.*

This results implies that

$$E [\hat{V}_0 (b, \mathbf{S}_0, \mathcal{N}_0, U_0)] \rightarrow B_0 (\mathbf{S}_0, \mathcal{N}_0, U_0) \quad (16)$$

as $b \rightarrow \infty$. Hence the estimator is asymptotically unbiased.

Theorem 5. (*Low estimator convergence*) Suppose that,

$$P \left[h_i \left(\mathbf{S}_i, \mathcal{N}_i, U_i, u^1 \right) + H_i \left(\mathbf{S}_i, \mathcal{N}_i - I_{\{u^1 \neq 0\}}, U_i + u^1 \right) \right. \\ \left. \neq h_i \left(\mathbf{S}_i, \mathcal{N}_i, U_i, u^2 \right) + H_i \left(\mathbf{S}_i, \mathcal{N}_i - I_{\{u^2 \neq 0\}}, U_i + u^2 \right) \right] = 1,$$

for $u^1, u^2 \in U_i, u^1 \neq u^2$ and all i . Then Theorem 4 also holds for the low estimator.

The additional condition imposed in Theorem 5 is analogous to that used in Theorem 3 of Broadie and Glasserman (1997). This condition says that, with probability one, the optimal exercise policy is never indifferent between the choices of volumes to exercise (including $u = 0$). As in Broadie and Glasserman (1997) imposing this condition simplifies the analysis of the estimator. Theorems 4 and 5 are proven in Appendix B.

2.4. Numerical Results

Pricing swing contracts involves a large number of parameters and in this section we provide some results which illustrate the validity of our method across a variety of specifications. We assume that the underlying assets follow a risk neutral stochastic process, there are no transaction costs and other than penalties, there are no other constraints considered. We also assume a constant risk free rate of interest and the volatilities of all assets are known constant functions of time.

The option swing rights may be exercised at discrete times up to and including expiry and the volume choices given are in discrete amounts. That is, anytime the holder chooses to exercise a right, they must choose from a finite list of possible volume amounts. The rationale behind allowing all time steps to be exercise opportunities is the exponential growth in computational time caused by adding intermediate non-exercise times. However, the method can easily be modified to incorporate these extra time steps. As previously mentioned penalties can be implemented globally and are based on the net volume exercised during the contract.

2.4.1. Single Dimension

Beginning with the one dimensional case, we have based our simulations on an underlying asset with a risk neutralized price process that satisfies the following stochastic differential equation,

$$dS_i = S_i [(r - \delta) dt + \sigma dZ_i]. \tag{17}$$

In this equation, r is the riskless interest rate, Z_i is a standard Brownian motion process, σ is a constant volatility parameter and the underlying asset itself pays a continuous dividend yield δ . The parameter values for the underlying asset are specified as $r = 0.05$, $\delta = 0.1$, and $\sigma = 0.2$.

The swing options considered have both up and down swing rights and the payoff upon exercise is

$$u \times \max [\max (S - K_u, 0), \max (K_d - S, 0), 0], \tag{18}$$

where u is the volume exercised, S is the price of the underlying asset at the exercise time, and K_u and K_d are the up and down swing strike prices, respectively. For the examples considered here, we set $K_u = K_d = K$ which simplifies the payoff function to

$$u \times \max (S - K, K - S). \tag{19}$$

For all examples in Section 2.4.1 the option expiry is 3.0 years and the options have both up and down swing rights with strike prices $K_u = K_d = 40.0$, respectively. In examples where the holder controls the amount exercised, a list of volume choices is given. The volume choices are consecutive integer multiples of a base amount and the up swing and down swing volumes have the same

magnitude. For comparison purposes the results in this subsection include a binomial value which is calculated using the Forest of Trees Lari et al. (2001).

All simulations in this subsection were completed on the SHARCNet cluster Whale. Whale is located at the University of Waterloo and consists of Opteron 2.2 GHz processors (four per node) with a Gigabit Ethernet interconnect. Timing results listed below are given in total cpu time accumulated which is approximately equal to (program runtime) × (number of processors used).

Example 1. (Illustration of Bias and Convergence) The swing option in this example has one up and one down swing right, three exercise opportunities, exercise volume of 60 units of the underlying and there is no penalty. The initial price is USD 40. Here we illustrate the effect of the branching factor b on the value estimates. Specifically, we perform R repeated valuations of the FOST with a branching factor of b and hold the total sample size fixed using the relation $R = 32,000 \left(\frac{10}{b}\right)$. Figure 2 plots the FOST estimates versus branching factor (the estimates are the averages of the R repeated valuations). Taking the Binomial estimate as the true option value, it is clear that the high estimator overestimates the true price while the low estimator underestimates the true price. Furthermore, as the branching factor increases, the high estimator decreases and the low estimator increases to the true option price, clearly illustrating estimator convergence. Estimator standard errors are approximately 0.07% of estimator value.

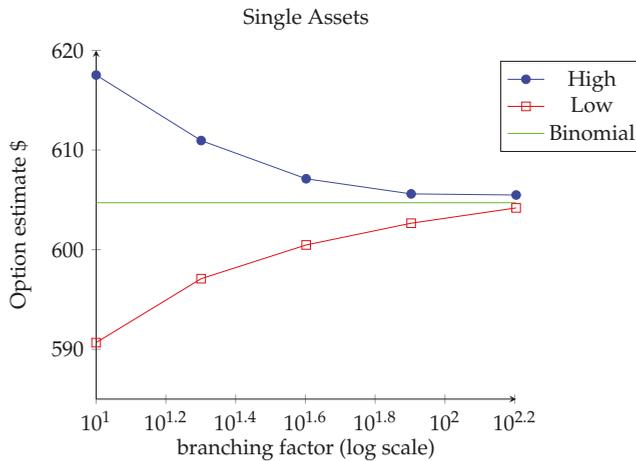


Figure 2. Option value estimates (USD) vs. log branching factor (b) with a single underlying asset. The option has one up and one down swing right, 3 exercise opportunities, exercise volume of 60 units of the underlying and there is no usage penalty. The initial price is USD 40. The number of repeated valuations $R = 32000 \left(\frac{10}{b}\right)$ results in standard errors $\approx 0.07\%$ of option value.

Example 2. (Effect of Usage Penalty and Initial Asset Price) In this example, the option has two up and two down swing rights and, upon exercise there are three volume choices—20, 40 and 60 units of the asset. Should the final net volume exercised exceed 90 units or be below -90 units a penalty is imposed. The penalty is calculated by multiplying the terminal asset price by ten times the excess usage above 90 units or below -90 units. To see the effect of the penalty on option value, we also turn the penalty term off and value the corresponding option with no penalty. The initial asset value ranges from USD 20 to USD 60 in steps of USD 10. There are $m = 5$ exercise opportunities and we use a branching factor of $b = 20$ with $R = 4000$ repeated valuations.

The pricing results are presented in Table 2. In each row of the table, we see that the high and low estimators bound the true price. Unsurprisingly, imposing a penalty on the cumulative volume reduces the option value. Furthermore, as the initial stock price increases, the increase in an up swing right’s value is more than the

decrease in a down swing right's value. The opposite is true as the initial stock price decreases. The end result is that as the initial stock price moves away from being at-the-money, the option value increases.

The average computing time per row (not including the binomial forest valuation) was 5.6 h for the cases with usage penalty and was 1.1 h without usage penalty. The reduction in runtime for the case with no penalty can be described as follows. If there are no constraints (e.g., penalty, storage) on the option holder then upon exercise it is always optimal to choose the maximum amount. Therefore with no penalty this option is equivalent to that of an otherwise identical swing option with no volume choices and an exercise volume of 60 units. The latter has fewer trees in its forest and is therefore quicker to evaluate. In Table 2 we have chosen to exploit this as a convenient way to save computational time. For the binomial method run times were on the order of a few seconds.

Table 2. Swing option values as a function of initial asset price and usage penalty with a single underlying asset. Parameter values used are $N_u = N_d = 2$, $U_i = \{20, 40, 60\}$, $b = 20$, $R = 4000$, $m = 5$, $U_{min} = -90$, and $U_{max} = 90$.

S_0	Penalty	High	Error	Low	Error	Binomial
60	ON	2271.153	1.418	2240.319	1.378	2259.845
	OFF	2422.781	1.576	2392.872	1.523	2411.844
50	ON	1445.468	0.844	1408.843	0.904	1429.645
	OFF	1542.053	0.978	1503.963	0.980	1526.055
40	ON	1018.104	0.859	968.793	1.044	989.651
	OFF	1156.591	0.911	1134.093	0.903	1145.801
30	ON	1345.556	1.205	1309.214	1.280	1326.266
	OFF	1562.347	1.316	1532.854	1.343	1546.055
20	ON	2189.531	0.905	2147.623	1.018	2157.976
	OFF	2443.877	0.924	2402.192	1.034	2412.354

Example 3. (Effect of Number of Exercise Rights) In this example we illustrate that the option value increases with the number of exercise rights and compare the swing option value with that of a corresponding basket of American options. The option has $m = 5$ exercise opportunities, an exercise volume of 60 units, and there is no usage penalty. Additionally, we set the initial price to $S_0 = 40$ and use a branching factor of $b = 20$ with $R = 4000$ repeated valuations. We consider options having an equal number of up and down swing rights. Table 3 gives the option price estimates for $N_u = N_d = 1, 3, 5$ along with prices computed using the Forest of Trees. First notice that the high and low estimates bound the true option from the binomial model. Next note that with $N_u = N_d = 5$ exercise rights, the high and low estimates are exactly the same. In this case the number of exercise opportunities is equal to the numbers of up and down swing rights and since the up and down swing strikes are equal, exactly one of these rights will be exercised at each opportunity (see Equation (19)). This makes both the exercise payoff and the continuation value estimates exactly the same at all times and along all branches for both the low and high estimators, yielding identical prices.

Second, the option value increases with the number of swing rights. However, the price increases by a factor that is less than the increase in the number of swing rights. For example, when the number rights increases by a factor of 3 (e.g., going from $N_u = N_d = 1$ to $N_u = N_d = 3$) the option value increases by a factor of 2.5 and when the number of rights increases by a factor of $\frac{5}{3}$ (e.g., going from $N_u = N_d = 3$ to $N_u = N_d = 5$), the option value increases by a factor of 1.2. This result matches with the intuition that a swing option with a given number of rights is less valuable than a basket of American put and call options with otherwise identical parameters and $K_d \leq K_u$.

Table 3. Swing option values as a function of the number of exercise rights with a single underlying asset. Parameter values used are exercise volume of 60 units, $S_0 = 40$, $b = 20$, $R = 4000$, $m = 5$, and no usage penalty.

$\mathcal{N}_u = \mathcal{N}_d$	High	Error	Low	Error	Binomial
1	630.054	0.449	605.394	0.453	617.832
3	1573.237	1.449	1559.517	1.437	1567.344
5	1852.788	2.128	1852.788	2.128	1852.627

For the case of one up and one down swing right the basket of American options would contain a single call corresponding to the up swing right and a single put corresponding to the down swing right, with equal strike prices for the call and put. Changing the number of up and down swing rights invokes a change in the corresponding number of call and put options in the basket. Figure 3 shows a comparison between the values of a basket of American options and a swing option with a comparable number of exercise rights. The value of the basket is linear in the number of exercise rights and the swing option value is below the value of the corresponding basket when the number of rights is greater than one. This follows from the restriction that only one swing right may be exercised at any exercise opportunity whereas all American options of a particular type could be exercised at a given time. In the case of one up and one down right the two are equal since it would never be optimal to exercise both the put and call style rights at the same time. As the number of rights increases, the difference in values increases due to the swing option restriction allowing only a single right to be exercised at each opportunity. The low-biased estimator is used for both the basket and swing option values in Figure 3.

Option Estimate vs. Exercise rights - Single Asset - Low Estimator

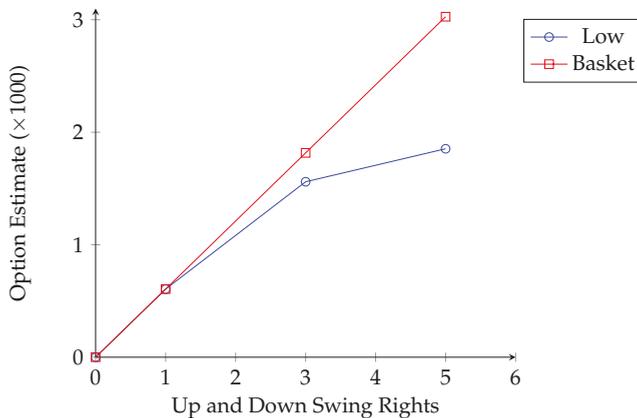


Figure 3. Basket of American calls and puts and swing option values versus the number of exercise rights using a single underlying asset. Parameter values used are exercise volume of 60 units, $S_0 = 40$, $b = 20$, $R = 4000$, $m = 5$, and no usage penalty.

2.4.2. Calibrated Forward Curve

In this example, we use the trinomial-tree model described in Section 4 of Jaiilet et al. (2004) from which price movements are simulated. This model is a 1-factor model with mean reversion that is seasonally adjusted and calibrated to the forward curve. The option we value is Example (a) of Section 4.2 in Jaiilet et al. (2004). This option a two up right swing option with each right allowing

the holder to take delivery of either 1 or 2 MMBTus of natural gas. We simplify to have four exercise opportunities and 4 months until expiry. Upon exercise the holder gets

$$\max (U_i (A_i S_i - K), 0) \tag{20}$$

where U_i is the volume chosen, A_i is the seasonality factor and S_i is the deseasonalized spot price.

Figure 4 plots the option price estimates, including 95% confidence intervals, against branching factor. The number of repeated valuations used to generate prices with a branching factor of b is $R = \frac{160,000}{b}$. We see that, with a branching factor of only 8, the confidence intervals for the high- and low-biased estimators begin to overlap and quickly become almost indistinguishable for higher branching factors, numerical illustration of both estimators' consistency. Additionally, the high and low estimators converge to the true price computed using the trinomial model. We note that the serial computational times (for a single FOST valuation) for branching factors of 8 and 32 were approximately 4.5 and 110 s, respectively using a 2.1 Ghz Core 2 Duo processor. The pricing results shown in Figure 4 are consistent with the results in [Jaillet et al. \(2004\)](#) but we note that the valuation method in that publication breaks down in higher dimensions and in cases where the inclusion of more risk factors is desirable.

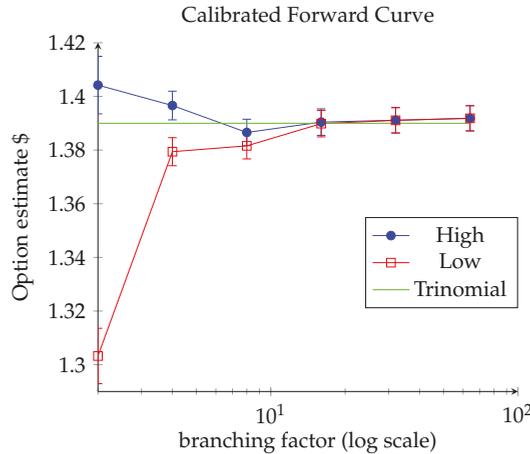


Figure 4. Option-value estimates versus (log) branching factor. Approximate pointwise 95% confidence intervals for each estimate are given by the vertical bars. The option and underlying model in this example is from [Jaillet et al. \(2004\)](#) with the Trinomial price given by their Forest of Trinomial Trees method.

2.4.3. Five Dimensions

Due to the computationally intensive nature of this method it only becomes truly useful in cases where PDE or tree based methods fail. In this subsection we present high-dimensional versions of the examples presented in Section 2.4.1. The asset prices are assumed to follow a risk neutralized correlated geometric Brownian motion described by the stochastic differential equations,

$$dS_i^k = S_i^k \left[(r - \delta^k) dt + \sigma^k dZ_i^k \right], \quad k = 1, \dots, d, \tag{21}$$

where Z_i^k is a standard Brownian motion process and the instantaneous correlation between Z^k and Z^s is ρ^{ks} . Here the parameter values are specified as $r = 0.05$, $\delta^k = \delta = 0.1$, $\sigma^k = \sigma = 0.2$ for all k and that $\rho^{ks} = 0$ for all $k \neq s$. In addition all assets have the same initial value, S_0 , and we take the number of assets to be $d = 5$.

The swing options considered have both up and down swing rights and the payoff upon exercise is

$$u \times \max \left[\left(\max_{k=1, \dots, d} S^k - K_u, K_d - \max_{k=1, \dots, d} S^k \right), 0 \right], \quad (22)$$

where S^k is the price of the k th underlying asset at the exercise time. This payoff is an extension of the example given in Broadie and Glasserman (1997) and Broadie and Glasserman (2004) for single-exercise American-style options. For the examples considered here, we set $K_u = K_d = K$ which simplifies the payoff function to

$$u \times \max \left(\max_{k=1, \dots, d} S^k - K, K - \max_{k=1, \dots, d} S^k \right). \quad (23)$$

As in Section 2.4.1, the option expiry is 3.0 years and the options have both up and down swing rights with strike prices $K_u = K_d = 40.0$, respectively. In examples where the holder controls the amount exercised, a list of volume choices is given. Note that we present results from our FOST methodology without comparisons to other methods as there is no generally accepted benchmark for the examples considered here.

Example 4. (Illustration of Bias and Convergence) This 5-dimensional example corresponds with Example 1. The swing option has one up and one down swing right, three exercise opportunities, exercise volume of 60 units of the underlying and there is no penalty. The initial price is USD 40. We perform R repeated valuations of the FOST with a branching factor of b and hold the total sample size fixed using the relation $R = 32000 \left(\frac{10}{b} \right)$. This results in standard errors $\approx 0.09\%$ of option value. Figure 5 plots the FOST estimates versus branching, showing that the high estimator overestimates the option price while the low estimator underestimates the price. Furthermore, as the branching factor increases, the high estimator decreases and the low estimator increases and they appear to be converging to the same value, illustrating estimator convergence. These findings are consistent with those in Example 1.

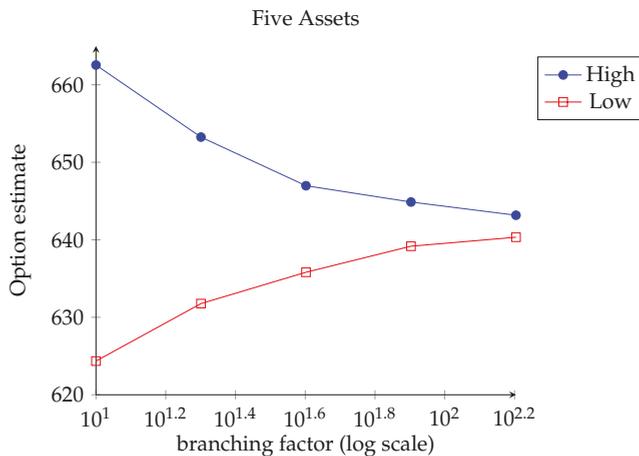


Figure 5. Option value estimates (USD) vs. log branching factor (b) with a five-dimensional underlying. The option has one up and one down swing right, three exercise opportunities, exercise volume of 60 units and there is no usage penalty. The number of repeated valuations $R = 32000 \left(\frac{10}{b} \right)$ results in standard errors $\approx 0.09\%$ of option value.

Example 5. (Effect of Usage Penalty and Initial Asset Price) This 5-dimensional example corresponds with Example 2, with the option specifications the same as presented there, modulo the adjustment to the payoff function to five dimensions. The pricing results are in Table 4. The effects of a usage penalty and initial stock

price are qualitatively the same compared with the 1-dimensional results. We note that the option value estimates in this example are higher than those in Example 2 due to the payoffs depending on the maximum of the five asset prices. The computing times for the 5-dimensional case are similar to those for the 1-dimensional asset.

Example 6. (Effect of Number of Exercise Rights) This 5-dimensional example corresponds with Example 3, with the option specifications the same as presented there, modulo the adjustment to the payoff function to five dimensions. The results are given in Table 5 and Figure 6. The results, intuition, and interpretation are qualitatively the same as the 1-dimensional results.

Table 4. Swing option values as a function of moneyness and penalties with a five-dimensional underlying asset. Parameter values used are $\mathcal{N}_u = \mathcal{N}_d = 2$, $\mathcal{U}_i = \{20, 40, 60\}$, $b = 20$, $R = 4000$, $m = 5$, $U_{min} = -90$, and $U_{max} = 90$.

S_0	Penalty	High	Error	Low	Error
60	ON	3577.280	2.864	3517.297	2.845
	OFF	3832.050	2.286	3772.123	2.856
50	ON	2246.657	2.280	2197.957	2.259
	OFF	2479.081	2.341	2431.065	2.318
40	ON	1221.847	1.595	1189.610	1.564
	OFF	1257.171	1.499	1226.370	1.467
30	ON	1105.831	0.453	1087.851	0.447
	OFF	1209.179	0.393	1196.255	0.391
20	ON	1937.615	0.445	1930.860	0.472
	OFF	2177.194	0.489	2177.031	0.513

Table 5. Swing option values as a function of the number of exercise rights with a five-dimensional underlying asset. Parameter values used are base volume = 60 units, $S_0 = 40$, $b = 20$, $R = 4000$, $m = 5$, and no usage penalty.

$\mathcal{N}_u = \mathcal{N}_d$	High	Error	Low	Error
1	683.144	0.741	652.481	0.721
3	1728.947	2.279	1709.497	2.248
5	2087.495	3.114	2087.495	3.114

Option Estimate vs. Exercise rights - Five Assets - Low Estimator

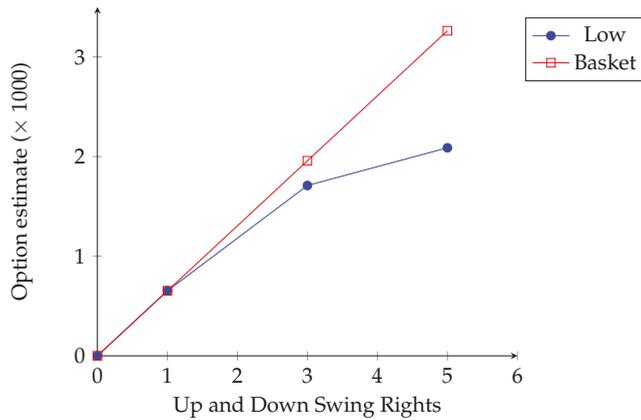


Figure 6. Basket of American calls and puts and swing option values versus the number of exercise rights using a 5-dimensional underlying asset. Parameter values used are exercise volume of 60 units, $S_0 = 40$, $b = 20$, $R = 4000$, $m = 5$, and no usage penalty.

2.5. Algorithmic Enhancement via Parallel Processing

One method for enhancing the computational efficiency of this algorithm is by taking advantage of multi-processor computing techniques. The simplest and most obvious implementation would be to parallelize across repeated valuations of the forest resulting in serial farming of the repeated valuations. Since each repeated valuation results in an iid random value for the option estimate, the generation of all the results may be completed independently of one another, removing the need for communication between processors. This method is simple and effective. However we state here without numerical evidence that it results in a near perfect speed up without the need for expensive interconnections. With this method the minimum run time that can be produced is determined by the number of processors available, the number of repeated valuations necessary for the desired accuracy and the run time of a single forest.

A variation on the aforementioned parallel implementation is to parallelize the FOST computations internally within the forest. In the results shown in Figure 7 the FOST algorithm has been modified so that the computation of the individual trees within the forest is done using multiple processors. Here we have begun the parallelization after the first time step by dividing up the computation of the remaining subtrees across different processors. Upon completion, the results are gathered and the option value at the initial time step is determined. In Figure 7 we see that this method results in a near perfect speed up due to the small ratio of communication time versus computational time. This implementation may be combined with serial farming resulting in further computational time efficiency. This is discussed more fully in Marshall et al. (2011).

In Figure 7 the swing option is identical one in Example 4 and pricing is done with a branching factor $b = 160$. The computational times were generated using the SHARCNET cluster Hound which comprises 2.2 GHz Opteron processors with 4 GB per core and Infini-Band interconnections. Run times are normalized to the run time of a single processor.

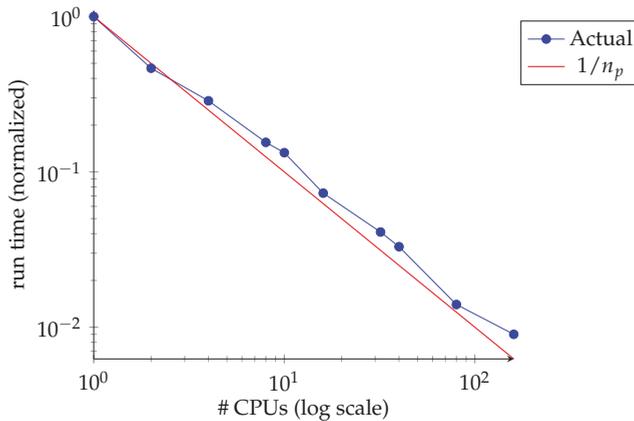


Figure 7. Normalized runtime using MPI versus number of CPUs (n_p). The option is identical to the no volume choice swing option with a five-dimensional underlying considered in Section 2.4.3. The branching factor used is $b = 160$.

3. Discussion and Conclusion

The FOST can be thought of as generalizations of two existing pricing methodologies. First, it generalizes the Forest of Trees method to a high-dimensional underlying. Second, it generalizes the Stochastic Tree pricing method for single-exercise right options to multiple exercise rights. We construct high and low FOST estimators analogous to those defined for the Stochastic Tree. We prove properties regarding FOST estimator bias, ordering, and convergence and present numerical results as illustrations.

In related work, we have replaced the binomial/trinomial trees in the Forest of Trees method with Stochastic Meshes [Broadie and Glasserman \(2004\)](#), creating the Forest of Stochastic Meshes [Marshall and Reesor \(2011\)](#). This avoids the exponential growth in computing time with the number of exercise opportunities experienced by the FOST. Another avenue of future work involves algorithmic enhancement. In Section 2.5 we discussed the use of parallel processing to reduce computing time. Two alternatives to this are variance reduction and bias reduction. There are some standard variance reduction methods (e.g., antithetic variates, control variates) that could be used to produce more efficient estimators. The bias reduction technique given in [Whitehead et al. \(2012\)](#) for Stochastic Tree estimators successfully reduces the branching factor required to obtain a desired accuracy for an American option value. This technique can be extended to correct the bias in FOST estimators and we have preliminary evidence of its effectiveness [Marshall \(2012\)](#). Combinations of variance reduction, bias reduction, and parallel processing can be investigated to further improve the algorithm’s performance.

Author Contributions: R.M.R. and T.J.M. contributed equally to all aspects of this paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded in part by the Natural Sciences and Engineering Research Council of Canada grant number RGPIN-2017-05441.

Acknowledgments: The authors thank SHARCNet for computational resources and technical support, particular acknowledgement goes to both Tyson Whitehead and Baolai Ge. We also thank graduate students and faculty in the Financial Mathematics group at Western University, in particular, Matt Davison, Adam Metzler, Rogemar Mamon, and Lars Stentoft.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results. The views represented herein are the authors’ own views and do not necessarily represent the views of Bank of Montreal or its affiliates and are not a product of its research.

Abbreviations

The following abbreviations are used in this manuscript:

MEO	Multiple Exercise Option
PDE	Partial Differential Equation
FOST	Forest of Stochastic Trees
FOSM	Forest of Stochastic Meshes
MC	Monte Carlo

Appendix A. Nomenclature

In this appendix we describe the notation used in this paper.

If X is a random variable, we write $\|X\|$ for the p -norm $(E|X|^p)^{1/p}$ of X . The conditional p -norm of X on \mathcal{Z}_i , $(E|X|^p|\mathcal{Z}_i)^{1/p}$, is denoted $\|X\|_{\mathcal{Z}_i}$. Here we include a summary of notation used for the proofs contained in this appendix as well as all subsequent appendices.

- Time is indexed by i for t_i , $i = 0, 1, \dots, m$.
- R is the number of repeated valuations of the forest.
- b is the branching factor.
- \mathbf{S}_i^j is the spot price vector at time t_i for branch $\mathbf{j} = \{j_0, j_1, \dots, j_i\}$. For convenience we may suppress the bold superscript if there is no ambiguity in doing so, in these cases \mathbf{S}_{i+1}^j refers to the time- t_{i+1} price along the branch path $\mathbf{j} = \{j_0, j_1, \dots, j_i, j\}$.
- \mathcal{Z}_i represents the time- t_i history of the set of state variables $(\mathbf{S}_i^j, \mathcal{N}_i, U_i)$, where we suppress the branching history index.
- $\hat{V}_i(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i)$ is the time- t_i , state- \mathcal{Z}_i high estimator.
- $\hat{v}_{il}(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i)$ is the time- t_i , state- \mathcal{Z}_i leave one out low biased estimator which does not include node l at time- t_{i+1} .
- $\hat{H}_{il}(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i, u)$ is the time- t_i , state- \mathcal{Z}_i leave one out hold value estimator for exercising u units which does not include node l at time- t_{i+1} ,

$$\hat{H}_{il}(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i, u) = \frac{1}{b-1} \sum_{\substack{k=1 \\ k \neq i}}^b D_{i+1} \hat{v}_{i+1}^k(b, \mathbf{S}_{i+1}^k, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u)$$

$$\mathbf{k} = \{j, k\}$$

- $\hat{g}_{il}(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i, u) = h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u) + \hat{H}_{il}(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i, u)$
- $\hat{v}_i(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i)$ is the time- t_i , state- \mathcal{Z}_i low estimator

$$\hat{v}_i = \frac{1}{b} \sum_{l=1}^b \hat{v}_{il}(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i)$$

- \mathcal{N}_i is the time- t_i number of exercise rights remaining.
- U_i is the time- t_i cumulative volume.
- \mathcal{U}_i is the time- t_i discretized set of available volume choices,

$$\mathcal{U}_i = \{u_0, u_1, u_2, \dots, u_z : z \in \mathbb{N}\},$$

where $u_0 = 0$.

- u is the time- t_i volume exercised. Here $u \in \mathcal{U}_i$.
- D_{i+1} is the discount factor from t_{i+1} to t_i .
- $h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u)$ is the time- t_i , state- \mathcal{Z}_i payoff from exercising u units with $h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, 0) = 0$.

- $H_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i)$ is the time- t_i , state- \mathcal{Z}_i true hold value,

$$H_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i) = \mathbb{E} \left[D_{i+1} B_{i+1}(\mathbf{S}_{i+1}^k, \mathcal{N}_{i+1}, U_{i+1}) \mid \mathcal{Z}_i \right]$$

- $B_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i)$ is the time- t_i , state- \mathcal{Z}_i true option value,

$$B_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i) = \max_{u \in \mathcal{U}_i} \left[h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u) + H_i(\mathbf{S}_i^j, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) \right]$$

where $I_{\{A\}}$ is the indicator function for set A .

Appendix B. Proofs of Main Results and Lemmas

Appendix B.1. Proofs of Main Results

In this section we prove the main results of the paper, Theorems 1–5, presented in Section 2.

Proof of Theorem 1. Here we prove the more general statement that $\mathbb{E}[\hat{V}_i(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i) \mid \mathcal{Z}_i] \geq B_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i)$ for $i = 0, 1, \dots, m$. The proof proceeds by backward induction. At expiry the inequality holds trivially since $\hat{V}_m(b, \mathbf{S}_m^j, \mathcal{N}_m, U_m) = B_m(\mathbf{S}_m^j, \mathcal{N}_m, U_m)$ so that $\mathbb{E}[\hat{V}_m(b, \mathbf{S}_m^j, \mathcal{N}_m, U_m) \mid \mathcal{Z}_m] \geq B_m(\mathbf{S}_m^j, \mathcal{N}_m, U_m)$. We now assume the inductive hypothesis, $\mathbb{E}[\hat{V}_{i+1}(b, \mathbf{S}_{i+1}^j, \mathcal{N}_{i+1}, U_{i+1}) \mid \mathcal{Z}_{i+1}] \geq B_{i+1}(\mathbf{S}_{i+1}^j, \mathcal{N}_{i+1}, U_{i+1})$, and proceed to the time- t_i case. We have,

$$\begin{aligned} & \mathbb{E} \left[\hat{V}_i(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i) \mid \mathcal{Z}_i \right] \\ &= \mathbb{E} \left[\max_{u \in \mathcal{U}_i} \left[h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u) + D_{i+1} \hat{V}_{i+1}(b, \mathbf{S}_{i+1}^k, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) \right] \mid \mathcal{Z}_i \right] \\ &\geq \max_{u \in \mathcal{U}_i} \left[h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u) + \mathbb{E} \left[D_{i+1} \hat{V}_{i+1}(b, \mathbf{S}_{i+1}^k, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) \mid \mathcal{Z}_i \right] \right] \\ &= \max_{u \in \mathcal{U}_i} \left[h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u) + \mathbb{E} \left[D_{i+1} \mathbb{E} \left[\hat{V}_{i+1}(b, \mathbf{S}_{i+1}^k, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) \mid \mathcal{Z}_{i+1} \right] \mid \mathcal{Z}_i \right] \right] \\ &\geq \max_{u \in \mathcal{U}_i} \left[h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u) + \mathbb{E} \left[D_{i+1} B_{i+1}(\mathbf{S}_{i+1}^k, \mathcal{N}_{i+1}, U_{i+1}) \mid \mathcal{Z}_i \right] \right] \\ &= \max_{u \in \mathcal{U}_i} \left[h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u) + H_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i) \right] \\ &= B_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i), \end{aligned}$$

where the first equality comes from the definition of the high estimator, the first inequality comes from the conditional Jensen’s inequality and note that $\mathcal{N}_{i+1} = \mathcal{N}_i - I_{\{u^* \neq 0\}}$ and $U_{i+1} = u_i + u^*$ where u^* is the value-maximizing volume choice, the second equality uses the tower law and the fact that D_{i+1} is \mathcal{Z}_i -measurable, and the second inequality invokes the inductive hypothesis. \square

Proof of Theorem 2. As with the proof of the bias of the high estimator we prove the more general statement that $\mathbb{E}[\hat{v}_i(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i) \mid \mathcal{Z}_i] \leq B_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i)$ for $i = 0, 1, \dots, m$ by backward induction. Again at expiry the inequality holds trivially since $\hat{v}_m(b, \mathbf{S}_m^j, \mathcal{N}_m, U_m) = B_m(\mathbf{S}_m^j, \mathcal{N}_m, U_m)$. We now assume the inductive hypothesis, $\mathbb{E}[\hat{v}_{i+1}(b, \mathbf{S}_{i+1}^j, \mathcal{N}_{i+1}, U_{i+1}) \mid \mathcal{Z}_{i+1}] \leq B_{i+1}(\mathbf{S}_{i+1}^j, \mathcal{N}_{i+1}, U_{i+1})$. We also note that since the $\hat{v}_{i\ell}$ ’s are iid we have that, $\mathbb{E}[\hat{v}_i \mid \mathcal{Z}_i] = \mathbb{E}[\hat{v}_{i\ell} \mid \mathcal{Z}_i]$. In what follows we define $\hat{u}_i^* \in \mathcal{U}_i$ to be the volume choice which maximizes a particular $\hat{v}_{i\ell}$. That is,

$$\hat{u}_i^* = \arg \max_{u \in \mathcal{U}_i} \left[\hat{g}_{i\ell}(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i, u) \right]. \tag{A1}$$

Note that $\hat{g}_{il}(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i, u)$ is conditionally independent of $\hat{v}_{i+1,l}(b, \mathbf{S}_{i+1}^1, \mathcal{N}_{i+1}, U_{i+1}, u)$ given \mathcal{Z}_i and subsequently \hat{u}_i^* is also independent of $\hat{v}_{i+1,l}$ given \mathcal{Z}_i since it is a function of \hat{g}_{il} .

Now,

$$\begin{aligned} \mathbb{E} \left[\hat{v}_{il} \left(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i \right) \mid \mathcal{Z}_i \right] &= \mathbb{E} \left[D_{i+1} \hat{v}_{i+1,l} \left(b, \mathbf{S}_{i+1}^k, \mathcal{N}_i, U_i \right) I_{\{\hat{u}_i^* = 0\}} \mid \mathcal{Z}_i \right] \\ &\quad + \mathbb{E} \left[\left(h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u_1 \right) + D_{i+1} \hat{v}_{i+1,l} \left(b, \mathbf{S}_{i+1}^k, \mathcal{N}_i - 1, U_i + u_1 \right) \right) I_{\{\hat{u}_i^* = u_1\}} \mid \mathcal{Z}_i \right] \\ &\quad + \dots + \mathbb{E} \left[\left(h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u_z \right) + D_{i+1} \hat{v}_{i+1,l} \left(i + 1, \mathbf{S}_{i+1}^k, \mathcal{N}_i - 1, U_i + u_z \right) \right) I_{\{\hat{u}_i^* = u_z\}} \mid \mathcal{Z}_i \right] \\ &= \mathbb{E} \left[D_{i+1} \hat{v}_{i+1,l} \left(b, \mathbf{S}_{i+1}^k, \mathcal{N}_i, U_i \right) \mid \mathcal{Z}_i \right] \mathbb{P}(\hat{u}_i^* = 0 \mid \mathcal{Z}_i) + h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u_1 \right) \mathbb{P}(\hat{u}_i^* = u_1 \mid \mathcal{Z}_i) \\ &\quad + \mathbb{E} \left[D_{i+1} \hat{v}_{i+1,l} \left(b, \mathbf{S}_{i+1}^k, \mathcal{N}_i - 1, U_i + u_1 \right) \mid \mathcal{Z}_i \right] \mathbb{P}(\hat{u}_i^* = u_1 \mid \mathcal{Z}_i) \\ &\quad + \dots + h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u_z \right) \mathbb{P}(\hat{u}_i^* = u_z \mid \mathcal{Z}_i) \\ &\quad + \mathbb{E} \left[D_{i+1} \hat{v}_{i+1,l} \left(b, \mathbf{S}_{i+1}^k, \mathcal{N}_i - 1, U_i + u_z \right) \mid \mathcal{Z}_i \right] \mathbb{P}(\hat{u}_i^* = u_z \mid \mathcal{Z}_i) \\ &= \mathbb{E} \left[D_{i+1} \hat{v}_{i+1,l} \left(b, \mathbf{S}_{i+1}^k, \mathcal{N}_i, U_i \right) \mid \mathcal{Z}_i \right] p_0 \\ &\quad + h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u_1 \right) p_1 + \mathbb{E} \left[D_{i+1} \hat{v}_{i+1,l} \left(b, \mathbf{S}_{i+1}^k, \mathcal{N}_i - 1, U_i + u_1 \right) \mid \mathcal{Z}_i \right] p_1 \\ &\quad + \dots + h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u_z \right) p_z + \mathbb{E} \left[D_{i+1} \hat{v}_{i+1,l} \left(b, \mathbf{S}_{i+1}^k, \mathcal{N}_i - 1, U_i + u_z \right) \mid \mathcal{Z}_i \right] p_z \end{aligned}$$

where in the second equality we have used the conditional independence of \hat{g}_{il} and $\hat{v}_{i+1,l}$. Here $p_0 = \mathbb{P}(\hat{u}_i^* = 0 \mid \mathcal{Z}_i)$ and $p_j = \mathbb{P}(\hat{u}_i^* = u_j \mid \mathcal{Z}_i)$ for $1 \leq j \leq z$ and $p_0 + \dots + p_z = 1$. Thus, using the tower law, we have,

$$\begin{aligned} \mathbb{E} \left[\hat{v}_i \left(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i \right) \mid \mathcal{Z}_i \right] &= \mathbb{E} \left[\hat{v}_{il} \left(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i \right) \mid \mathcal{Z}_i \right] \\ &= \mathbb{E} \left[D_{i+1} \mathbb{E} \left[\hat{v}_{i+1,l} \left(b, \mathbf{S}_{i+1}^k, \mathcal{N}_i, U_i \right) \mid \mathcal{Z}_{i+1} \right] \mid \mathcal{Z}_i \right] p_0 \\ &\quad + h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u_1 \right) p_1 + \mathbb{E} \left[D_{i+1} \mathbb{E} \left[\hat{v}_{i+1,l} \left(b, \mathbf{S}_{i+1}^k, \mathcal{N}_i - 1, U_i + u_1 \right) \mid \mathcal{Z}_{i+1} \right] \mid \mathcal{Z}_i \right] p_1 \\ &\quad + \dots + h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u_z \right) p_z + \mathbb{E} \left[D_{i+1} \mathbb{E} \left[\hat{v}_{i+1,l} \left(b, \mathbf{S}_{i+1}^k, \mathcal{N}_i - 1, U_i + u_z \right) \mid \mathcal{Z}_{i+1} \right] \mid \mathcal{Z}_i \right] p_z \\ &\leq \mathbb{E} \left[D_{i+1} B_{i+1} \left(\mathbf{S}_{i+1}^k, \mathcal{N}_i, U_i \right) \mid \mathcal{Z}_i \right] p_0 \\ &\quad + h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u_1 \right) p_1 + \mathbb{E} \left[D_{i+1} B_{i+1} \left(\mathbf{S}_{i+1}^k, \mathcal{N}_i - 1, U_i + u_1 \right) \mid \mathcal{Z}_i \right] p_1 \\ &\quad + \dots + h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u_z \right) p_z + \mathbb{E} \left[D_{i+1} B_{i+1} \left(\mathbf{S}_{i+1}^k, \mathcal{N}_i - 1, U_i + u_z \right) \mid \mathcal{Z}_i \right] p_z \\ &= H_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i \right) p_0 + \left(h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u_1 \right) + H_i \left(\mathbf{S}_i^j, \mathcal{N}_i - 1, U_i + u_1 \right) \right) p_1 \\ &\quad + \dots + \left(h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u_z \right) + H_i \left(\mathbf{S}_i^j, \mathcal{N}_i - 1, U_i + u_z \right) \right) p_z \\ &\leq \max_{u \in \mathcal{U}_i} \left[h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u \right) + H_i \left(\mathbf{S}_i^j, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u \right) \right] \\ &= B_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i \right) \end{aligned}$$

where the first inequality follows from the inductive hypothesis and the remaining steps follow from the definitions for B_i and H_i . \square

Proof of Theorem 3. At expiry we have that $\hat{v}_m(b, \mathbf{S}_m^j, \mathcal{N}_m, U_m) = \hat{V}_m(b, \mathbf{S}_m^j, \mathcal{N}_m, U_m) = B_m(\mathbf{S}_m^j, \mathcal{N}_m, U_m)$ so the relation holds trivially. We now take the inductive hypothesis to be

$\hat{v}_{i+1}(b, \mathbf{S}_{i+1}^j, \mathcal{N}_{i+1}, U_{i+1}) \leq \hat{V}_{i+1}(b, \mathbf{S}_{i+1}^j, \mathcal{N}_{i+1}, U_{i+1})$ for $j_{i+1} = 1, \dots, b$. Using the \hat{g}_{il} as defined above we first consider the case where for a given tree,

$$\hat{u}_l^* = \arg \max_{u \in \mathcal{U}_l} [\hat{g}_{il}(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i, u)], \tag{A2}$$

is the same for all l (i.e., $\hat{u}_l^* = \hat{u}^*$, for all l).

Then,

$$\begin{aligned} \hat{v}_i(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i) &= \frac{1}{b} \sum_{l=1}^b \hat{v}_{il}(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i) \\ &= \frac{1}{b} \sum_{l=1}^b [h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^*) + D_{i+1} \hat{v}_{i+1}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{\hat{u}^* \neq 0\}}, U_i + \hat{u}^*)] \\ &\leq \frac{1}{b} \sum_{l=1}^b [h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^*) + D_{i+1} \hat{V}_{i+1}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{\hat{u}^* \neq 0\}}, U_i + \hat{u}^*)] \\ &= h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^*) + \frac{1}{b} \sum_{l=1}^b [D_{i+1} \hat{V}_{i+1}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{\hat{u}^* \neq 0\}}, U_i + \hat{u}^*)] \\ &\leq \max_{u \in \mathcal{U}_i} \left[h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u) + \frac{1}{b} \sum_{l=1}^b [D_{i+1} \hat{V}_{i+1}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u)] \right] \\ &= \hat{V}_i(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i) \end{aligned}$$

where the first inequality comes from the inductive hypothesis and the remaining relations come from the parameter definitions.

Next consider the case where the low estimator gives two different estimated optimal exercise amounts, \hat{u}^1, \hat{u}^2 , across all l branches where $\hat{u}^1 \neq \hat{u}^2$. That is $\hat{u}_l^* = \hat{u}^1$ or $\hat{u}_l^* = \hat{u}^2$ for all $l = 1, \dots, b$. As above we take \hat{u}_l^* to be the optimal exercise amount determined by the l -th leave one out estimator, then,

$$\begin{aligned} \hat{v}_i(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i) &= \frac{1}{b} \sum_{l=1}^b \hat{v}_{il}(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i) \\ &= \frac{1}{b} \sum_{l=1}^b [h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}_l^*) + D_{i+1} \hat{v}_{i+1}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{\hat{u}_l^* \neq 0\}}, U_i + \hat{u}_l^*)] \\ &= \frac{1}{b} \sum_{l=1}^b \left\{ [h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^1) + D_{i+1} \hat{v}_{i+1}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{\hat{u}^1 \neq 0\}}, U_i + \hat{u}^1)] I_{\{\hat{u}_l^* = \hat{u}^1\}} \right. \\ &\quad \left. + [h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^2) + D_{i+1} \hat{v}_{i+1}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{\hat{u}^2 \neq 0\}}, U_i + \hat{u}^2)] I_{\{\hat{u}_l^* = \hat{u}^2\}} \right\} \\ &= \frac{\left(\frac{1}{b} \sum_{l=1}^b I_{\{\hat{u}_l^* = \hat{u}^1\}} \right) \times \left(\frac{1}{b} \sum_{l=1}^b [h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^1) + D_{i+1} \hat{v}_{i+1}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{\hat{u}^1 \neq 0\}}, U_i + \hat{u}^1)] I_{\{\hat{u}_l^* = \hat{u}^1\}} \right)}{\frac{1}{b} \sum_{l=1}^b I_{\{\hat{u}_l^* = \hat{u}^1\}}} \\ &\quad + \frac{\left(\frac{1}{b} \sum_{l=1}^b I_{\{\hat{u}_l^* = \hat{u}^2\}} \right) \times \left(\frac{1}{b} \sum_{l=1}^b [h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^2) + D_{i+1} \hat{v}_{i+1}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{\hat{u}^2 \neq 0\}}, U_i + \hat{u}^2)] I_{\{\hat{u}_l^* = \hat{u}^2\}} \right)}{\frac{1}{b} \sum_{l=1}^b I_{\{\hat{u}_l^* = \hat{u}^2\}}} \\ &= p \times \frac{\frac{1}{b} \sum_{l=1}^b [h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^1) + D_{i+1} \hat{v}_{i+1}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{\hat{u}^1 \neq 0\}}, U_i + \hat{u}^1)] I_{\{\hat{u}_l^* = \hat{u}^1\}}}{\frac{1}{b} \sum_{l=1}^b I_{\{\hat{u}_l^* = \hat{u}^1\}}} \\ &\quad + (1-p) \times \frac{\frac{1}{b} \sum_{l=1}^b [h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^2) + D_{i+1} \hat{v}_{i+1}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{\hat{u}^2 \neq 0\}}, U_i + \hat{u}^2)] I_{\{\hat{u}_l^* = \hat{u}^2\}}}{\frac{1}{b} \sum_{l=1}^b I_{\{\hat{u}_l^* = \hat{u}^2\}}}, \end{aligned}$$

where $p = \frac{1}{b} \sum_{l=1}^b I_{\{\hat{u}_l^* = \hat{u}^1\}}$.

Without loss of generality, suppose that $\hat{u}_l^* = \hat{u}^1$ for $l = 1, \dots, k$ and $\hat{u}_l^* = \hat{u}^2$ for $l = k + 1, \dots, b$. Then the above ratios become

$$\frac{\sum_{l=1}^k [h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^1) + D_{i+1} \hat{v}_{i+1}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{\hat{u}^1 \neq 0\}}, U_i + \hat{u}^1)]}{k} \tag{A3}$$

and

$$\frac{\sum_{l=k+1}^b \left[h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^2 \right) + D_{i+1} \hat{\vartheta}_{i+1} \left(b, \mathbf{S}_{i+1}^1, \mathcal{N}_i - I_{\{\hat{u}^2 \neq 0\}}, U_i + \hat{u}^2 \right) \right]}{b-k}, \tag{A4}$$

respectively. Now for any $i^* \leq k < j^* \leq b$ we have

$$\hat{g}_{\hat{u}^*} \left(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^1 \right) > \hat{g}_{ij^*} \left(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^1 \right)$$

which from the definition of \hat{g}_{il} , implies that

$$D_{i+1} \hat{\vartheta}_{i+1} \left(b, \mathbf{S}_{i+1}^{i^*}, \mathcal{N}_i, U_i + \hat{u}^1 \right) \leq D_{i+1} \hat{\vartheta}_{i+1} \left(b, \mathbf{S}_{i+1}^{j^*}, \mathcal{N}_i, U_i + \hat{u}^1 \right).$$

Therefore,

$$\max_{1 \leq a \leq k} \left[D_{i+1} \hat{\vartheta}_{i+1} \left(b, \mathbf{S}_{i+1}^a, \mathcal{N}_i, U_i + \hat{u}^1 \right) \right] \leq \min_{k+1 \leq a \leq b} \left[D_{i+1} \hat{\vartheta}_{i+1} \left(b, \mathbf{S}_{i+1}^a, \mathcal{N}_i, U_i + \hat{u}^1 \right) \right].$$

This implies that Equation (A3)

$$\begin{aligned} & \frac{1}{k} \sum_{l=1}^k \left[h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^1 \right) + D_{i+1} \hat{\vartheta}_{i+1} \left(b, \mathbf{S}_{i+1}^1, \mathcal{N}_i - I_{\{\hat{u}^1 \neq 0\}}, U_i + \hat{u}^1 \right) \right] \\ & \leq \frac{1}{b} \sum_{l=1}^b \left[h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^1 \right) + D_{i+1} \hat{\vartheta}_{i+1} \left(b, \mathbf{S}_{i+1}^1, \mathcal{N}_i - I_{\{\hat{u}^1 \neq 0\}}, U_i + \hat{u}^1 \right) \right], \end{aligned}$$

and similarly for Equation (A4)

$$\begin{aligned} & \frac{1}{b-k} \sum_{l=k+1}^b \left[h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^2 \right) + D_{i+1} \hat{\vartheta}_{i+1} \left(b, \mathbf{S}_{i+1}^1, \mathcal{N}_i - I_{\{\hat{u}^2 \neq 0\}}, U_i + \hat{u}^2 \right) \right] \\ & \leq \frac{1}{b} \sum_{l=1}^b \left[h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^2 \right) + D_{i+1} \hat{\vartheta}_{i+1} \left(b, \mathbf{S}_{i+1}^1, \mathcal{N}_i - I_{\{\hat{u}^2 \neq 0\}}, U_i + \hat{u}^2 \right) \right]. \end{aligned}$$

Therefore

$$\begin{aligned} \hat{\vartheta}_i \left(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i \right) & \leq p \times \frac{1}{b} \sum_{l=1}^b \left[h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^1 \right) + D_{i+1} \hat{\vartheta}_{i+1} \left(b, \mathbf{S}_{i+1}^1, \mathcal{N}_i - I_{\{\hat{u}^1 \neq 0\}}, U_i + \hat{u}^1 \right) \right] \\ & \quad + (1-p) \times \frac{1}{b} \sum_{l=1}^b \left[h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^2 \right) + D_{i+1} \hat{\vartheta}_{i+1} \left(b, \mathbf{S}_{i+1}^1, \mathcal{N}_i - I_{\{\hat{u}^2 \neq 0\}}, U_i + \hat{u}^2 \right) \right] \\ & \leq p \times \frac{1}{b} \sum_{l=1}^b \left[h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^1 \right) + D_{i+1} \hat{V}_{i+1} \left(b, \mathbf{S}_{i+1}^1, \mathcal{N}_i - I_{\{\hat{u}^1 \neq 0\}}, U_i + \hat{u}^1 \right) \right] \\ & \quad + (1-p) \times \frac{1}{b} \sum_{l=1}^b \left[h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^2 \right) + D_{i+1} \hat{V}_{i+1} \left(b, \mathbf{S}_{i+1}^1, \mathcal{N}_i - I_{\{\hat{u}^2 \neq 0\}}, U_i + \hat{u}^2 \right) \right] \\ & \leq \max \left[\frac{1}{b} \sum_{l=1}^b \left[h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^1 \right) + D_{i+1} \hat{V}_{i+1} \left(b, \mathbf{S}_{i+1}^1, \mathcal{N}_i - I_{\{\hat{u}^1 \neq 0\}}, U_i + \hat{u}^1 \right) \right], \right. \\ & \quad \left. \frac{1}{b} \sum_{l=1}^b \left[h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, \hat{u}^2 \right) + D_{i+1} \hat{V}_{i+1} \left(b, \mathbf{S}_{i+1}^1, \mathcal{N}_i - I_{\{\hat{u}^2 \neq 0\}}, U_i + \hat{u}^2 \right) \right] \right] \\ & \leq \max_{u \in \hat{\mathcal{U}}_i} \left[\frac{1}{b} \sum_{l=1}^b \left[h_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u \right) + D_{i+1} \hat{V}_{i+1} \left(b, \mathbf{S}_{i+1}^1, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u \right) \right] \right] \\ & = \hat{V}_i \left(\mathbf{S}_i^j, \mathcal{N}_i, U_i \right), \end{aligned}$$

where the second inequality comes from the inductive hypothesis, the third inequality is an application of Jensen's inequality, the fourth inequality comes from maximizing over a larger set, and the final equality is the definition of the high-biased estimator.

For the cases where the low estimator gives z^* distinct estimated optimal exercise amounts, $\hat{u}^1, \dots, \hat{u}^{z^*}$, across all z branches, $z^* = 3, \dots, z$, arguments similar to those given above (for 2 distinct estimated optimal exercise amounts) show that,

$$\hat{v}_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i) \leq \hat{V}_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i).$$

Since we restrict the number of volume choices to be finite, the theorem is proven. \square

Prior to proving Theorems 4 and 5 we first state the following preliminary result.

Lemma A1. *If $\|h_i(\mathbf{S}_i, \mathcal{N}_i, U_i, u)\| < \infty$ for all t_i , for some $p \geq 1$, then the following are true for all $0 \leq t_i \leq t_k \leq t_m$:*

$$\|B_k(\mathbf{S}_k, \mathcal{N}_k, U_k)\|_{\mathcal{Z}_i} < \infty \tag{A5}$$

$$\sup_b \|\hat{V}_k(b, \mathbf{S}_k, \mathcal{N}_k, U_k)\|_{\mathcal{Z}_i} < \infty \tag{A6}$$

$$\sup_b \|\hat{v}_k(b, \mathbf{S}_k, \mathcal{N}_k, U_k)\|_{\mathcal{Z}_i} < \infty \tag{A7}$$

The proof of this lemma can be found in the appendix. A second preliminary result is

Lemma A2. *Let $a_1, \dots, a_n, b_1, \dots, b_n, c_1, \dots, c_n$ be real numbers. Then,*

$$A_n \equiv |\max(a_1 + b_1, \dots, a_n + b_n) - \max(a_1 + c_1, \dots, a_n + c_n)| \leq 2 \sum_{i=1}^{n+1} |b_i - c_i| \equiv B_n. \tag{A8}$$

Its proof can also be found in the appendix. We now prove Theorems 4 and 5.

Proof of Theorem 4. Here we take $R = 1$ and state that if the convergence holds for a single realization of the forest then it will hold for the mean of any number of realizations due to the independence of each repeated valuation. Here we prove by backward induction the more general statement $\|\hat{V}_i(b, \mathbf{S}_i, \mathcal{N}_i, U_i) - B_i(\mathbf{S}_i, \mathcal{N}_i, U_i)\|_{\mathcal{Z}_i} \rightarrow 0$ for any generic node in a given tree and for all $i = 0, \dots, m$. At expiry the relation holds trivially since at $t_i = t_m$ we have that $\hat{V}_m(b, \mathbf{S}_m, \mathcal{N}_m, U_m) = B_m(\mathbf{S}_m, \mathcal{N}_m, U_m)$. The inductive hypothesis is taken to be $\|\hat{V}_{i+1}(b, \mathbf{S}_{i+1}, \mathcal{N}_{i+1}, U_{i+1}) - B_{i+1}(\mathbf{S}_{i+1}, \mathcal{N}_{i+1}, U_{i+1})\|_{\mathcal{Z}_{i+1}} \rightarrow 0$.

Now,

$$\begin{aligned}
 & \left\| \hat{V}_i(b, \mathbf{S}_i, \mathcal{N}_i, U_i) - B_i(\mathbf{S}_i, \mathcal{N}_i, U_i) \right\|_{\mathcal{Z}_i} \\
 &= \left\| \max_{u \in \mathcal{U}_i} \left[h_i(\mathbf{S}_i, \mathcal{N}_i, U_i, u) + \frac{1}{b} \sum_{j=1}^b D_{i+1} \hat{V}_{i+1}(b, \mathbf{S}_{i+1}^j, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) \right] \right. \\
 &\quad \left. - \max_{u \in \mathcal{U}_i} \left[h_i(\mathbf{S}_i, \mathcal{N}_i, U_i, u) + H_i(\mathbf{S}_i, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) \right] \right\|_{\mathcal{Z}_i} \\
 &= \left\| \max_{u \in \mathcal{U}_i} \left[h_i(\mathbf{S}_i, \mathcal{N}_i, U_i, u) + \frac{1}{b} \sum_{j=1}^b D_{i+1} \hat{V}_{i+1}(b, \mathbf{S}_{i+1}^j, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) \right] \right. \\
 &\quad \left. - \max_{u \in \mathcal{U}_i} \left[h_i(\mathbf{S}_i, \mathcal{N}_i, U_i, u) + H_i(b, \mathbf{S}_i^j, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) \right] \right\|_{\mathcal{Z}_i} \\
 &\leq \left\| 2 \sum_{k=0}^z \frac{1}{b} \sum_{j=1}^b D_{i+1} \hat{V}_{i+1}(b, \mathbf{S}_{i+1}^j, \mathcal{N}_i - I_{\{u_k \neq 0\}}, U_i + u_k) - H_i(\mathbf{S}_i, \mathcal{N}_i - I_{\{u_k \neq 0\}}, U_i + u_k) \right\|_{\mathcal{Z}_i} \\
 &\leq 2 \sum_{k=0}^z \left\| \frac{1}{b} \sum_{j=1}^b D_{i+1} \hat{V}_{i+1}(b, \mathbf{S}_{i+1}^j, \mathcal{N}_i - I_{\{u_k \neq 0\}}, U_i + u_k) - H_i(\mathbf{S}_i, \mathcal{N}_i - I_{\{u_k \neq 0\}}, U_i + u_k) \right\|_{\mathcal{Z}_i} \\
 &\leq 2 \sum_{k=0}^z \left\| \frac{1}{b} \sum_{j=1}^b D_{i+1} [\hat{V}_{i+1}(i, \mathbf{S}_{i+1}^j, \mathcal{N}_i - I_{\{u_k \neq 0\}}, U_i + u_k) - B_{i+1}(\mathbf{S}_{i+1}^j, \mathcal{N}_i - I_{\{u_k \neq 0\}}, U_i + u_k)] \right\|_{\mathcal{Z}_i} \\
 &\quad + 2 \sum_{k=0}^z \left\| D_{i+1} B_{i+1}(\mathbf{S}_{i+1}^j, \mathcal{N}_i - I_{\{u_k \neq 0\}}, U_i + u_k) - H_i(\mathbf{S}_i, \mathcal{N}_i - I_{\{u_k \neq 0\}}, U_i + u_k) \right\|_{\mathcal{Z}_i} \\
 &= 2 \sum_{k=0}^z (E_k + C_k)
 \end{aligned}$$

where the first equality comes from the definitions of the estimator and the true value. The third step comes as a result of Lemma A2, the fourth step comes from a generalization of the triangle inequality. In the final step we rewrite the expression for convenience in what follows.

First we deal with the C_k 's. Given \mathcal{Z}_i we have that $D_{i+1} B_{i+1}(\mathbf{S}_{i+1}^j, \mathcal{N}_i - I_{\{u_k \neq 0\}}, U_i + u_k)$ for $j = 1, \dots, b$ and $k = 0, \dots, z$ are iid with means of $H_i(\mathbf{S}_i, \mathcal{N}_i - I_{\{u_k \neq 0\}}, U_i + u_k)$ and finite p -norms. Then by Theorem I.4.1 of Gut (1988) we have that all C_k 's in the above expression go to zero.

Next we consider the E_k 's. Here we have, by the properties of p -norms and the fact that the terms being averaged are iid, that

$$E_k \leq \left\| \hat{V}_{i+1}(b, \mathbf{S}_{i+1}, \mathcal{N}_i - I_{\{u_k \neq 0\}}, U_i + u_k) - B_{i+1}(\mathbf{S}_{i+1}, \mathcal{N}_i - I_{\{u_k \neq 0\}}, U_i + u_k) \right\|_{\mathcal{Z}_i}$$

since E_k is bounded by the p -norm of any one of the terms being averaged. By the inductive hypothesis

$$\left\| \hat{V}_{i+1}(b, \mathbf{S}_{i+1}, \mathcal{N}_{i+1}, U_{i+1}) - B_{i+1}(\mathbf{S}_{i+1}, \mathcal{N}_{i+1}, U_{i+1}) \right\|_{\mathcal{Z}_{i+1}} \rightarrow 0,$$

where $\mathcal{N}_{i+1} = \mathcal{N}_i - I_{\{u_k \neq 0\}}$ and $U_{i+1} = U_i + u_k$.

Also by a standard condition for uniform integrability (see Gut (1988) p. 178) we have that

$$\left\| \hat{V}_{i+1}(b, \mathbf{S}_{i+1}, \mathcal{N}_{i+1}, U_{i+1}) - B_{i+1}(\mathbf{S}_{i+1}, \mathcal{N}_{i+1}, U_{i+1}) \right\|_{\mathcal{Z}_i} \rightarrow 0, \tag{A9}$$

provided

$$\sup_b E \left[\left| \hat{V}_{i+1}(b, \mathbf{S}_{i+1}, \mathcal{N}_{i+1}, U_{i+1}) - B_{i+1}(\mathbf{S}_{i+1}, \mathcal{N}_{i+1}, U_{i+1}) \right|^{p+\epsilon} \mid \mathcal{Z}_i \right] < \infty$$

for some ϵ . From Lemma A1 we know that

$$\sup_b E \left[|\hat{V}_{i+1}(b, \mathbf{S}_{i+1}, \mathcal{N}_{i+1}, U_{i+1})|^{p+\epsilon} | \mathcal{Z}_i \right] < \infty$$

and that

$$E \left[|B_{i+1}(\mathbf{S}_{i+1}, \mathcal{N}_{i+1}, U_{i+1})|^{p+\epsilon} | \mathcal{Z}_i \right] < \infty.$$

Thus (A9) holds for each $k = 0, \dots, z$ and hence the result is proven. \square

Proof of Theorem 5. As with the proof of Theorem 4 we proceed by backward induction. Again at expiry the relation holds trivially since $\hat{v}_m(b, \mathbf{S}_m, \mathcal{N}_m, U_m) = B_m(\mathbf{S}_m, \mathcal{N}_i, U_i)$. The inductive hypothesis is taken to be $\|\hat{v}_{i+1}(b, \mathbf{S}_{i+1}, \mathcal{N}_{i+1}, U_{i+1}) - B_{i+1}(\mathbf{S}_{i+1}, \mathcal{N}_{i+1}, U_{i+1})\|_{\mathcal{Z}_{i+1}} \rightarrow 0$.

Let $\hat{g}_{il}(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i, u)$ be as defined at the start of Appendix A and note that, with probability one,

$$\begin{aligned} & h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u^1) + H_i(\mathbf{S}_i^j, \mathcal{N}_i - I_{\{u^1 \neq 0\}}, U_i + u^1) \\ & \neq h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u^2) + H_i(\mathbf{S}_i^j, \mathcal{N}_i - I_{\{u^2 \neq 0\}}, U_i + u^2), \end{aligned}$$

for all $u^1, u^2 \in \mathcal{U}_i, u^1 \neq u^2$.

Before proceeding, we stop to make three claims:

1. $\left\| \frac{1}{b} \sum_{l=1}^b D_{i+1} \hat{v}_{i+1}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) - H_i(\mathbf{S}_i^j, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) \right\|_{\mathcal{Z}_i} \rightarrow 0$
2. $\left\| \hat{g}_{il}(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i, u) - \left[h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u) - H_i(\mathbf{S}_i^j, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) \right] \right\|_{\mathcal{Z}_i} \rightarrow 0$
3. $\left\| I_{\{\hat{u}_i^* = u\}} - I_{\{u^* = u\}} \right\|_{\mathcal{Z}_i} \rightarrow 0$

for all $u \in \mathcal{U}_i$ and where

$$\begin{aligned} \hat{u}_i^* &= \arg \max_{u \in \mathcal{U}_i} \left[\hat{g}_{il}(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i, u) \right] \quad \text{and} \\ u^* &= \arg \max_{u \in \mathcal{U}_i} \left[h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u) - H_i(\mathbf{S}_i^j, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) \right] \end{aligned}$$

The proof of item (i) is the same as the proof of the corresponding step in Theorem 4. Since the estimators in (i) and (ii) differ only in the omission of one term in \hat{g}_{il} , similar arguments prove that (ii) also holds.

Now for (iii), if $\hat{u}_i^* = u^*$ then the result holds trivially. Now suppose that $\hat{u}_i^* = u \neq u^*$ for some $u \in \mathcal{U}_i$. Then,

$$\begin{aligned} & \left\| I_{\{\hat{u}_i^* = u\}} - I_{\{u^* = u\}} \right\|_{\mathcal{Z}_i} = \left\| I_{\{\hat{u}_i^* = u\}} \right\|_{\mathcal{Z}_i} = [\mathbb{P}(\hat{u}_i^* = u | \mathcal{Z}_i)]^{\frac{1}{p}} \\ & = \left[\mathbb{P} \left(\hat{g}_{il}(b, \mathbf{S}_i^j, \mathcal{N}_i, U_i, u) \geq h_i(\mathbf{S}_i^j, \mathcal{N}_i, U_i, u) + H_i(\mathbf{S}_i^j, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) \right) \right]^{1/p} \\ & \rightarrow 0. \end{aligned}$$

Since (ii) holds and convergence in p -norm implies convergence in probability. Thus (iii) is proven.

Now proceeding from the definition of the low estimator and the true option value for all $u \in \mathcal{U}_i$,

$$\begin{aligned}
 & \|\hat{v}_i(b, \mathbf{S}_i, \mathcal{N}_i, U_i) - B_i(\mathbf{S}_i, \mathcal{N}_i, U_i)\|_{\mathcal{Z}_i} \\
 &= \left\| \frac{1}{b} \sum_{l=1}^b \hat{v}_{il}(b, \mathbf{S}_i, \mathcal{N}_i, U_i) - B_i(\mathbf{S}_i, \mathcal{N}_i, U_i) \right\|_{\mathcal{Z}_i} \\
 &= \left\| \frac{1}{b} \sum_{l=1}^b \left(D_{i+1} \hat{v}_{i+1,l}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i, U_i) I_{\{\hat{a}_i^* = 0\}} \right. \right. \\
 &\quad \left. \left. + \left[ah_i(\mathbf{S}_i, \mathcal{N}_i, U_i, u_1) + D_{i+1} \hat{v}_{i+1,l}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - 1, U_i + u_1) \right] I_{\{\hat{a}_i^* = u_1\}} \right. \right. \\
 &\quad \left. \left. + \dots + \left[ah_i(\mathbf{S}_i, \mathcal{N}_i, U_i, u_z) + D_{i+1} \hat{v}_{i+1,l}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - 1, U_i + u_z) \right] I_{\{\hat{a}_i^* = u_z\}} \right) - B_i(\mathbf{S}_i, \mathcal{N}_i, U_i) \right\|_{\mathcal{Z}_i} \\
 &\leq \left\| \frac{1}{b} \sum_{l=1}^b D_{i+1} \hat{v}_{i+1,l}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i, U_i) I_{\{\hat{a}_i^* = 0\}} - H_i(\mathbf{S}_i, \mathcal{N}_i, U_i) I_{\{u^* = 0\}} \right\|_{\mathcal{Z}_i} \\
 &\quad + \left\| \frac{1}{b} \sum_{l=1}^b D_{i+1} \hat{v}_{i+1,l}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - 1, U_i + u_1) I_{\{\hat{a}_i^* = u_1\}} - H_i(\mathbf{S}_i, \mathcal{N}_i - 1, U_i + u_1) I_{\{u^* = u_1\}} \right\|_{\mathcal{Z}_i} \\
 &\quad + \dots + \left\| \frac{1}{b} \sum_{l=1}^b D_{i+1} \hat{v}_{i+1,l}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - 1, U_i + u_z) I_{\{\hat{a}_i^* = u_z\}} - H_i(\mathbf{S}_i, \mathcal{N}_i - 1, U_i + u_z) I_{\{u^* = u_z\}} \right\|_{\mathcal{Z}_i} \\
 &\quad + \left\| h_i(\mathbf{S}_i, \mathcal{N}_i, U_i, u_1) I_{\{\hat{a}_i^* = u_1\}} - h_i(\mathbf{S}_i, \mathcal{N}_i, U_i, u_1) I_{\{u^* = u_1\}} \right\|_{\mathcal{Z}_i} \\
 &\quad + \dots + \left\| h_i(\mathbf{S}_i, \mathcal{N}_i, U_i, u_z) I_{\{\hat{a}_i^* = u_z\}} - h_i(\mathbf{S}_i, \mathcal{N}_i, U_i, u_z) I_{\{u^* = u_z\}} \right\|_{\mathcal{Z}_i}
 \end{aligned} \tag{A10}$$

where the inequality in the third step is due to a generalization of the triangle inequality.

The immediate consequence of claim (iii) above is that all terms in Equation (A10) with the form

$$\left\| h_i(\mathbf{S}_i, \mathcal{N}_i, U_i, u) I_{\{\hat{a}_i^* = u\}} - h_i(\mathbf{S}_i, \mathcal{N}_i, U_i, u) I_{\{u^* = u\}} \right\|_{\mathcal{Z}_i} \rightarrow 0$$

for all $u \in \mathcal{U}_i$. Thus

$$\sum_{k=0}^z \left\| h_i(\mathbf{S}_i, \mathcal{N}_i, U_i, u_k) I_{\{\hat{a}_i^* = u_k\}} - h_i(\mathbf{S}_i, \mathcal{N}_i, U_i, u_k) I_{\{u^* = u_k\}} \right\|_{\mathcal{Z}_i} \rightarrow 0.$$

It remains to show that the remaining terms in (A10) converge in the p -norm. Taking one of these terms, that is, fix a $u \in \mathcal{U}_i$, we now show this converges in the p -norm to zero.

$$\begin{aligned}
 & \left\| \frac{1}{b} \sum_{l=1}^b D_{i+1} \hat{v}_{i+1,l}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) I_{\{\hat{a}_i^* = u\}} - H_i(\mathbf{S}_i, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) I_{\{u^* = u\}} \right\|_{\mathcal{Z}_i} \\
 &= \left\| \frac{1}{b} \sum_{l=1}^b \left[D_{i+1} \hat{v}_{i+1,l}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) I_{\{\hat{a}_i^* = u\}} - [D_{i+1} \hat{v}_{i+1,l}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) I_{\{u^* = u\}}] \right. \right. \\
 &\quad \left. \left. + \frac{1}{b} \sum_{l=1}^b D_{i+1} \hat{v}_{i+1,l}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) I_{\{u^* = u\}} - H_i(\mathbf{S}_i, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) I_{\{u^* = u\}} \right] \right\|_{\mathcal{Z}_i} \\
 &\leq \left\| \frac{1}{b} \sum_{l=1}^b \left[D_{i+1} \hat{v}_{i+1,l}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) I_{\{\hat{a}_i^* = u\}} - D_{i+1} \hat{v}_{i+1,l}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) I_{\{u^* = u\}} \right] \right\|_{\mathcal{Z}_i} \\
 &\quad + \left\| \frac{1}{b} \sum_{l=1}^b D_{i+1} \hat{v}_{i+1,l}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) I_{\{u^* = u\}} - H_i(\mathbf{S}_i, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) I_{\{u^* = u\}} \right\|_{\mathcal{Z}_i} \\
 &\leq \left\| D_{i+1} \hat{v}_{i+1,l}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) \right\|_{\mathcal{Z}_i} \cdot \left\| I_{\{\hat{a}_i^* = u\}} - I_{\{u^* = u\}} \right\|_{\mathcal{Z}_i} \\
 &\quad + \left\| I_{\{u^* = u\}} \right\|_{\mathcal{Z}_i} \cdot \left\| \frac{1}{b} \sum_{l=1}^b D_{i+1} \hat{v}_{i+1,l}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) - H_i(\mathbf{S}_i, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) \right\|_{\mathcal{Z}_i},
 \end{aligned}$$

where the first step comes from adding and subtracting the same term, the second comes from applying the triangle inequality and the third step comes from factoring out common terms.

Now by (iii),

$$\left\| I_{\{\hat{a}_i^* = u\}} - I_{\{u^* = u\}} \right\|_{\mathcal{Z}_i} \rightarrow 0,$$

by (i),

$$\left\| \frac{1}{b} \sum_{l=1}^b D_{i+1} \hat{v}_{i+1,l}(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) - H_i(\mathbf{S}_i, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u) \right\|_{\mathcal{Z}_i} \rightarrow 0,$$

and we note that

$$\begin{aligned} & \left\| I_{\{u^* = u\}} \right\|_{\mathcal{Z}_i} < \infty \quad \text{and} \\ & \left\| D_{i+1} \hat{v}_{i+1,l} \left(b, \mathbf{S}_{i+1}^l, \mathcal{N}_i - I_{\{u \neq 0\}}, U_i + u \right) \right\|_{\mathcal{Z}_i} < \infty, \end{aligned}$$

by (A7).

Hence we have proven the consistency of the low-biased estimator. \square

Appendix B.2. Lemma Proofs

Proof of Lemma A1. If every $h_i(\mathbf{S}_i, \mathcal{N}_i, U_i, u)$ has finite p -th moment, then each $\|h_i(\mathbf{S}_k, \mathcal{N}_k, U_k, u)\|_{\mathcal{Z}_i}$ is finite. Since the max, discounting, and conditional expectation operators preserve finiteness of moments then it follows that $\|B_k(\mathbf{S}_k, \mathcal{N}_k, U_k)\|_{\mathcal{Z}_i}$ and also $\|H_k(\mathbf{S}_k, \mathcal{N}_k, U_k)\|_{\mathcal{Z}_i}$ must also be finite.

Proceeding to (A6), fix t_i and proceed by backward induction on t_k from t_m to t_i . At expiry (A6) follows from (A5). Then for $t_k < t_m$,

$$\begin{aligned} & \sup_b \left\| \hat{V}_k(b, \mathbf{S}_k, \mathcal{N}_k, U_k) \right\|_{\mathcal{Z}_i} = \sup_b \left\| \max_{u \in \mathcal{U}_k} \left[h_k(\mathbf{S}_k, \mathcal{N}_k, U_k, u) + \hat{H}_k(b, \mathbf{S}_k, \mathcal{N}_k - I_{\{u \neq 0\}}, U_k + u) \right] \right\|_{\mathcal{Z}_i} \\ & = \sup_b \left\| \max_{u \in \mathcal{U}_k, u} \left[h_k(\mathbf{S}_k, \mathcal{N}_k, U_k, u) + \frac{1}{b} \sum_{j=1}^b D_{k+1} \hat{V}_{k+1}(b, \mathbf{S}_{k+1}^j, \mathcal{N}_k - I_{\{u \neq 0\}}, U_k + u) \right] \right\|_{\mathcal{Z}_i} \\ & \leq \|h_k(\mathbf{S}_k, \mathcal{N}_k, U_k, 0)\|_{\mathcal{Z}_i} + \sup_b \left\| \frac{1}{b} \sum_{j=1}^b D_{k+1} \hat{V}_{k+1}(b, \mathbf{S}_{k+1}^j, \mathcal{N}_k, U_k) \right\|_{\mathcal{Z}_i} \\ & \quad + \|h_k(\mathbf{S}_k, \mathcal{N}_k, U_k, u_1)\|_{\mathcal{Z}_i} + \sup_b \left\| \frac{1}{b} \sum_{j=1}^b D_{k+1} \hat{V}_{k+1}(b, \mathbf{S}_{k+1}^j, \mathcal{N}_k - 1, U_k + u_1) \right\|_{\mathcal{Z}_i} \\ & \quad + \dots + \|h_k(\mathbf{S}_k, \mathcal{N}_k, U_k, u_r)\|_{\mathcal{Z}_i} + \sup_b \left\| \frac{1}{b} \sum_{j=1}^b D_{k+1} \hat{V}_{k+1}(b, \mathbf{S}_{k+1}^j, \mathcal{N}_k - 1, U_k + u_r) \right\|_{\mathcal{Z}_i} \\ & \leq \sup_b \left\| \hat{V}_{k+1}(b, \mathbf{S}_{k+1}, \mathcal{N}_k, U_k) \right\|_{\mathcal{Z}_i} + \|h_k(\mathbf{S}_k, \mathcal{N}_k, U_k, u_1)\|_{\mathcal{Z}_i} + \sup_b \left\| \hat{V}_{k+1}(b, \mathbf{S}_{k+1}, \mathcal{N}_k - 1, U_k + u_1) \right\|_{\mathcal{Z}_i} \\ & \quad + \dots + \|h_k(\mathbf{S}_{k+1}, \mathcal{N}_k, U_k, u_r)\|_{\mathcal{Z}_i} + \sup_b \left\| \hat{V}_{k+1}(b, \mathbf{S}_{k+1}, \mathcal{N}_k - 1, U_k + u_r) \right\|_{\mathcal{Z}_i}, \end{aligned}$$

where $h_k(\mathbf{S}_k, \mathcal{N}_k, U_k, 0) = 0$. This is the sum of a finite number of terms, each of which is finite.

For (iii) the proof is similar to that of (ii). \square

Proof of Lemma A2. In order to prove Lemma A2 we proceed by induction by considering the cases for $n = 1$ and $n = 2$. For $n = 1$,

$$A_1 = |\max(a_1 + b_1) - \max(a_1 + c_1)| = |b_1 - c_1| < B_1$$

therefore $A_1 \leq B_1$. Now, for $n = 2$

$$A_1 = |\max(a_1 + b_1, a_2 + b_2) - \max(a_1 + c_1, a_2 + c_2)|$$

Consider the following,

1. $a_1 + b_1 > a_2 + b_2$

(a) $a_1 + c_1 > a_2 + c_2$

Then

$$A_2 = |a_1 + b_1 - a_1 - c_1| = |b_1 - c_1| \leq B_2$$

(b) $a_1 + c_1 < a_2 + c_2$

Note that conditions (i) and (b) imply that

$$b_2 - b_1 < a_1 - a_2 < c_2 - c_1 \tag{A11}$$

and we have

$$\begin{aligned} A_2 &= |a_1 + b_1 - a_2 - c_2| \\ &= |a_1 + b_1 - c_1 + c_1 - a_2 - c_2| \\ &\leq |b_1 - c_1| + |(a_1 - a_2) - (c_2 - c_1)| \\ &\leq |b_1 - c_1| + |(b_2 - b_1) - (c_2 - c_1)| \\ &= |b_1 - c_1| + |b_2 - c_2 - (b_1 - c_1)| \\ &\leq 2|b_1 - c_1| + |b_2 - c_2| \\ &\leq 2|b_1 - c_1| + 2|b_2 - c_2| \\ &= B_2 \end{aligned}$$

where the first inequality comes from the triangle inequality, the second comes from Inequality (A11) and the third inequality comes from another application of the triangle inequality.

2. $a_2 + b_2 > a_1 + b_1$

(a) $a_2 + c_2 > a_1 + c_1$

Then

$$A_2 = |a_2 + b_2 - a_2 - c_2| = |b_2 - c_2| \leq B_2$$

(b) $a_2 + c_2 < a_1 + c_1$

Note that conditions (ii) and (b) imply that

$$b_1 - b_2 < a_2 - a_1 < c_1 - c_2 \tag{A12}$$

and we have

$$\begin{aligned} A_2 &= |a_2 + b_2 - a_1 - c_1| \\ &= |a_2 + b_2 - c_2 + c_2 - a_1 - c_1| \\ &\leq |b_2 - c_2| + |(a_2 - a_1) - (c_1 - c_2)| \\ &\leq |b_2 - c_2| + |(b_1 - b_2) - (c_1 - c_2)| \\ &= |b_2 - c_2| + |b_1 - c_1 - (b_2 - c_2)| \\ &\leq 2|b_2 - c_2| + |b_1 - c_1| \\ &\leq 2|b_2 - c_2| + 2|b_1 - c_1| \\ &= B_2 \end{aligned}$$

where again the first inequality comes from the triangle inequality, the second comes from Inequality (A12) and the third inequality comes from another application of the triangle inequality.

Therefore $A_2 \leq B_2$.

Now assume that the inductive hypothesis $A_n \leq B_n$ is true. We need to show that $A_{n+1} \leq B_{n+1}$. First define i_n and j_n such that

$$a_{i_n} + b_{i_n} = \max(a_1 + b_1, \dots, a_n + b_n)$$

and

$$a_{i_n} + c_{i_n} = \max(a_1 + c_1, \dots, a_n + c_n)$$

respectively. Now,

$$\begin{aligned} A_{n+1} &= |\max(a_1 + b_1, \dots, a_n + b_n, a_{n+1} + b_{n+1}) - \max(a_1 + c_1, \dots, a_n + c_n, a_{n+1} + c_{n+1})| \\ &= |\max(a_{i_n} + b_{i_n}, a_{n+1} + b_{n+1}) - \max(a_{j_n} + c_{j_n}, a_{n+1} + c_{n+1})| \end{aligned}$$

Consider the following,

1. $a_{i_n} + b_{i_n} > a_{n+1} + b_{n+1}$

(a) $a_{j_n} + c_{j_n} > a_{n+1} + c_{n+1}$

$$\begin{aligned} A_{n+1} &= |a_{i_n} + b_{i_n} - a_{j_n} - c_{j_n}| \\ &\leq 2 \sum_{i=1}^n |b_i - c_i| \\ &\leq 2 \sum_{i=1}^{n+1} |b_i - c_i| \\ &= B_{n+1} \end{aligned}$$

where the first inequality comes from the inductive hypothesis.

(b) $a_{j_n} + c_{j_n} < a_{n+1} + c_{n+1}$

By the definitions of i_n and j_n and (b) we have

$$a_{i_n} + c_{i_n} \leq a_{j_n} + c_{j_n} < a_{n+1} + c_{n+1}.$$

This combined with (i) gives

$$b_{n+1} - b_{i_n} < a_{i_n} - a_{n+1} < c_{n+1} - c_{i_n} \tag{A13}$$

Then

$$\begin{aligned} A_{n+1} &= |a_{i_n} + b_{i_n} - a_{n+1} - c_{n+1}| \\ &= |a_{i_n} + b_{i_n} - c_{i_n} + c_{i_n} - a_{n+1} - c_{n+1}| \\ &\leq |b_{i_n} - c_{i_n}| + |(a_{i_n} - a_{n+1}) - (c_{n+1} - c_{i_n})| \\ &\leq |b_{i_n} - c_{i_n}| + |(b_{n+1} - b_{i_n}) - (c_{n+1} - c_{i_n})| \\ &= |b_{i_n} - c_{i_n}| + |b_{n+1} - c_{n+1} - (b_{i_n} - c_{i_n})| \\ &\leq 2|b_{i_n} - c_{i_n}| + |b_{n+1} - c_{n+1}| \\ &\leq B_{n+1} \end{aligned}$$

where again the first inequality comes from the triangle inequality, the second comes from Inequality (A13) and the third inequality comes from another application of the triangle inequality.

2. $a_{i_n} + b_{i_n} < a_{n+1} + b_{n+1}$

(a) $a_{j_n} + c_{j_n} < a_{n+1} + c_{n+1}$

$$\begin{aligned} A_{n+1} &= |a_{n+1} + b_{n+1} - a_{n+1} - c_{n+1}| \\ &= |b_{n+1} - c_{n+1}| \\ &\leq 2 \sum_{i=1}^{n+1} |b_i - c_i| \\ &= B_{n+1} \end{aligned}$$

(b) $a_{j_n} + c_{j_n} > a_{n+1} + c_{n+1}$

By the definitions of i_n and j_n and (ii) we have

$$a_{j_n} + b_{j_n} \leq a_{i_n} + b_{i_n} < a_{n+1} + b_{n+1}.$$

This combined with (b) gives

$$b_{j_n} - b_{n+1} < a_{n+1} - a_{j_n} < c_{j_n} - c_{n+1} \tag{A14}$$

Then

$$\begin{aligned} A_{n+1} &= |a_{n+1} + b_{n+1} - a_{j_n} - c_{j_n}| \\ &= |a_{n+1} + b_{n+1} - c_{n+1} + c_{n+1} - a_{j_n} - c_{j_n}| \\ &\leq |b_{n+1} - c_{n+1}| + |(a_{n+1} - a_{j_n}) - (c_{j_n} - c_{n+1})| \\ &\leq |b_{n+1} - c_{n+1}| + |(b_{j_n} - b_{n+1}) - (c_{j_n} - c_{n+1})| \\ &= |b_{n+1} - c_{n+1}| + |b_{j_n} - c_{j_n} - (b_{n+1} - c_{n+1})| \\ &\leq 2|b_{n+1} - c_{n+1}| + |b_{j_n} - c_{j_n}| \\ &\leq B_{n+1} \end{aligned}$$

where again the first inequality comes from the triangle inequality, the second comes from Inequality (A14) and the third inequality comes from another application of the triangle inequality.

Therefore $A_{n+1} \leq B_{n+1}$ and the Lemma is proven. \square

References

Andersen, Leif, and Mark Broadie. 2004. Primal-Dual Simulation Algorithm for Pricing Multidimensional American Options. *Management Science* 50: 1222–34. [CrossRef]

Bally, Vlad, Gilles Pagés, and Jacques Printems. 2005. A Quantization Tree Method for Pricing and Hedging Multidimensional American Options. *Mathematical Finance* 15: 119–68. [CrossRef]

Barraquand, Jerome, and Didier Martineau. 1995. Numerical Valuation of High Dimensional Multivariate American Securities. *Journal of Financial and Quantitative Analysis* 30: 383–405. [CrossRef]

Barrera-Esteve, Christophe, Florent Bergeret, Charles Dossal, Emmanuel Gobet, Asma Meziou, Remi Munos, and Damien Reboul-Salze. 2006. Numerical Methods for the Pricing of Swing Options: A Stochastic Control Approach. *Methodology and Computing in Applied Probability* 8: 517–40. [CrossRef]

Ben Latifa, Imene, Joseph Frederic Bonnans, and Mohamed Mnif. 2016. Numerical methods for an optimal multiple stopping problem. *Stochastics and Dynamics* 16.

Bender, Christian. 2011. Dual pricing of multi-exercise options under volume constraints. *Finance and Stochastics* 15: 1–26. [CrossRef]

Bender, Christian, and John Schoenmakers. 2006. An Iterative Algorithm for Multiple Stopping: Convergence and Stability. *Advances in Applied Probability* 38: 729–49. [CrossRef]

- Broadie, Mark, and Paul Glasserman. 1997. Pricing American-style securities using simulation. *Journal of Economic Dynamics and Control* 21: 1323–52. [\[CrossRef\]](#)
- Broadie, Mark, and Paul Glasserman. 2004. A stochastic mesh method for pricing high-dimensional American options. *The Journal of Computational Finance* 7: 35–72. [\[CrossRef\]](#)
- Calvo-Garrido, Maria del Carmen, Matthias Ehrhardt, and Carlos Vázquez. 2017. Pricing swing options in electricity markets with two stochastic factors using a partial differential equation approach. *The Journal of Computational Finance* 20: 81–107. [\[CrossRef\]](#)
- Carriere, Jacques. 1996. Valuation of Early-Exercise Price of Options Using Simulations and Nonparametric Regression. *Insurance: Mathematics and Economics* 19: 19–30.
- Chandramouli, Shyam S., and Martin B. Haugh. 2012. A unified approach to multiple stopping and duality. *Operations Research Letters* 40: 258–64. [\[CrossRef\]](#)
- Chen, Zhuliang, and Peter A. Forsyth. 2007. A semi-Lagrangian approach for natural gas storage valuation and optimal operation. *SIAM Journal on Scientific Computing* 30: 339–68. [\[CrossRef\]](#)
- Clément, Emmanuelle, Damien Lamberton, and Philip Protter. 2001. An analysis of a least-squares regression algorithm for American option pricing. *Finance Stochastics* 6: 449–71. [\[CrossRef\]](#)
- Cox, John C., Stephen A. Ross, and Mark Rubinstein. 1979. Option Pricing: A Simplified Approach. *Journal of Financial Economics* 7: 229–63. [\[CrossRef\]](#)
- Deng, Shijie, and Shmuel S. Oren. 2006. Electricity derivatives and risk management. *Energy* 31: 940–53. [\[CrossRef\]](#)
- Gut, Allan. 1988. *Stopped Random Walks*. New York: Springer.
- Gyurkó, Lajos Gergely, Ben M. Hambly, and Jan Hendrick Witte. 2015. Monte Carlo methods via a dual approach for some discrete time stochastic control problems. *Mathematical Methods of Operations Research* 81: 109–35. [\[CrossRef\]](#)
- Haugh, Martin B., and Leonid Kogan. 2004. Pricing American Options: A Duality Approach. *Operations Research* 52: 258–70. [\[CrossRef\]](#)
- Ibáñez, Alfredo. 1996. Valuation by Simulation of Contingent Claims with Multiple Early Exercise Opportunities. *Mathematical Finance* 19: 19–30.
- Jaillet, Patrick, Ehud I. Ronn, and Stathis Tompaidis. 2004. Valuation of Commodity-Based Swing Options. *Management Science* 50: 909–21. [\[CrossRef\]](#)
- Kim, Dong-Hyun, Eul-Bum Lee, In-Hyeo Jung, and Douglas Alleman. 2019. The Efficacy of the Tolling Model's Ability to Improve Project Profitability on International Steel Plants. *Energies* 12: 1221. [\[CrossRef\]](#)
- Dahlgren, Martin, and Ralf Korn. 2005. The Swing Option on the Stock Market. *The International Journal of Theoretical and Applied Finance* 8: 123–129. [\[CrossRef\]](#)
- Lari-Lavassani, Ali, Mohamadreza Simchi, and Antony Ware. 2001. A Discrete Valuation of Swing Options. *Canadian Applied Mathematics Quarterly* 9: 35–73.
- Longstaff, Francis A., and Eduardo S. Schwartz. 2001. Valuing American Options by Simulation: A Simple Least-squares Approach. *The Review of Financial Studies* 14: 113–47. [\[CrossRef\]](#)
- Ludkovski, Michael, and Rene Carmona. 2010. Valuation of energy storage: An optimal switching approach. *Quantitative Finance* 10: 359–74.
- Marshall, T. James. 2012. Valuation of Multiple Exercise Options. Ph.D. thesis, Western University, London, ON, Canada.
- Marshall, T. James, and R. Mark Reesor. 2011. Forest of Stochastic Meshes: A Method for Valuing High Dimensional Swing Options. *Operations Research Letters* 39: 17–21. [\[CrossRef\]](#)
- Marshall, T. James, R. Mark Reesor, and Matthew Cox. 2011. Simulation Valuation of Multiple Exercise Options. Paper presented at the 2011 Winter Simulation Conference, Phoenix, AZ, USA, December 11–14.
- Meinshausen, Nicolai, and Ben M. Hambly. 2004. Monte Carlo Methods For the Valuation of Multiple-Exercise Options. *Mathematical Finance* 14: 557–83. [\[CrossRef\]](#)
- Stentoft, Lars. 2004. Assessing the Least-Squares Monte-Carlo Approach to American Option Valuation. *Review of Derivatives Research* 7: 129–68. [\[CrossRef\]](#)
- Thompson, Matt, Matt Davison, and Henning Rasmussen. 2009. Natural gas storage valuation and optimization: A real options application. *Naval Research Logistics* 56: 226–38. [\[CrossRef\]](#)
- Tilley, James A. 1993. Valuing American Options in a Path Simulation Model. *Transactions of the Society of Actuaries* 45: 83–104.

Whitehead, Tyson, R. Mark Reesor, and Matt Davison. 2012. A bias-reduction technique for Monte Carlo pricing of early-exercise options. *Journal of Computational Finance* 15. [[CrossRef](#)]

Wilhelm, Martina, and Christoph Winter. 2008. Finite element valuation of swing options. *Journal of Computational Finance* 11: 107–32. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Article

Bootstrapping the Early Exercise Boundary in the Least-Squares Monte Carlo Method

Pascal Létourneau ¹ and Lars Stentoft ^{2,*}

¹ Department of Finance and Business Law, University of Wisconsin-Whitewater, Whitewater, WI 53190, USA; letournp@uww.edu

² Department of Economics and Department of Statistical and Actuarial Sciences, University of Western Ontario, London, ON N6A 5C2, Canada

* Correspondence: lars.stentoft@uwo.ca

Received: 15 November 2019; Accepted: 11 December 2019; Published: 15 December 2019

Abstract: This paper proposes an innovative algorithm that significantly improves on the approximation of the optimal early exercise boundary obtained with simulation based methods for American option pricing. The method works by exploiting and leveraging the information in multiple cross-sectional regressions to the fullest by averaging the individually obtained estimates at each early exercise step, starting from just before maturity, in the backwards induction algorithm. With this method, less errors are accumulated, and as a result of this, the price estimate is essentially unbiased even for long maturity options. Numerical results demonstrate the improvements from our method and show that these are robust to the choice of simulation setup, the characteristics of the option, and the dimensionality of the problem. Finally, because our method naturally disassociates the estimation of the optimal early exercise boundary from the pricing of the option, significant efficiency gains can be obtained by using less simulated paths and repetitions to estimate the optimal early exercise boundary than with the regular method.

Keywords: American options; least-squares Monte Carlo; exercise boundary; simulation

JEL Classification: C15; G12; G13

1. Introduction

Simulation and regression based methods are by now standard methods used for American option pricing. Simulation methods are flexible, work in multiple dimensions, can be used with even the most complicated models and payoff functions, and are straightforward to use. Essentially, if you can simulate it (the underlying dynamics), you can price it (the derivatives contract).¹ Moreover, under standard assumptions, the estimates have nice properties since averages over independent samples converge to expected values. Almost all the simulation methods that have been proposed for American option pricing share the feature that in one way or another, they involve estimating or approximating the early exercise boundary or, more generally, estimating or approximating the conditional expectations that are used to decide whether or not to exercise at each step.² In the Least-Squares Monte Carlo (LSM) method of Longstaff and Schwartz (2001), for example, the conditional expectations are approximated with polynomials, and based on these approximations,

¹ See also Boyer and Stentoft (2013) for applications of this method to longevity risk products in insurance.

² See Ibanez and Zapatero (2004) for an example of a simulation method that directly approximates the optimal early exercise boundary.

it is then decided at each early exercise point and along each simulated path whether the option should be exercised or not; implicitly determining the early exercise region.

While the asymptotic properties of the LSM method are well established (see for example (Stentoft 2004b)), the quality of the price estimate obtained with this method depends on the order of the polynomial, M , used in the cross-sectional regression and the number of paths, N , used in the simulation (see, e.g., Moreno and Navas 2003; Stentoft 2004a). With a finite choice of polynomial terms, M , and simulated paths, N , the early exercise boundary is estimated with some error. Because of this, sub-optimal decisions are made, and everything else equal, low biased price estimates are obtained. Moreover, because sub-optimal decisions are made early on, errors are propagated backwards through the algorithm. This paper presents an innovative bias reduction technique directly applicable to the LSM method, as well as to many other similar methods. The method successfully delivers at each early exercise step, starting from just before maturity, a “better” estimated early exercise boundary. Because of this, less errors are accumulated, and by recursively correcting the estimated optimal early exercise boundary, we end up with essentially unbiased price estimates even for long maturity options.

To motivate our methodology, we first demonstrate that significantly less biased estimates can be obtained if the precision of the early exercise boundary is improved upon by simply averaging it over several independently repeated simulations. Next, we show that by averaging at each step, and therefore obtaining a less noisy estimate of the early exercise strategy, less errors are propagated backwards through the algorithm. Our method is based on the simple observation that if one uses the average of the approximated early exercise strategies instead of, say, the $I = 100$ individually determined strategies, the estimated prices are much closer to the benchmark prices. Our proposed method exploits and leverages this observation to the fullest by averaging at each step in the backwards induction algorithm instead of only after having gone through all the steps of the backward induction algorithm. We refer to our method as a bootstrapping approach because of the similarities it has with how the term structure of interest rates is bootstrapped and the way bootstrapping is used for inference in regression models. As a result of the bootstrapping approach, the final bias of the estimated price is almost completely eliminated.

We compare the results from our bootstrapping approach to the regular method of Longstaff and Schwartz (2001) and show that our proposed method works extremely well across various levels of moneyness, maturity, and volatility. Moreover, compared to the regular LSM method, the bootstrapping method is much less sensitive to the choice of polynomial order used in the cross-sectional regressions and to the number of paths used in the Monte Carlo simulation. When a reasonable order, say $M = 9$, is used in the regressions and when a standard number of paths, say $N = 100,000$, is used in the simulation, our proposed method essentially provides unbiased price estimates. We also demonstrate that the technique generalizes straightforwardly to multi-asset options with various payoffs. To illustrate this, we price arithmetic and geometric average options and options on the maximum and minimum of three assets. For all four payoff functions, the results hold, and the bootstrapped method always performs the best and delivers price estimates with negligible or very small bias for reasonable choices of the number of regressors and paths used in the Monte Carlo simulation.

Our method successfully corrects the bias even in very small samples and allows for using, e.g., $N = 10,000$ paths or less to determine the early exercise boundary. Our detailed results also show that one does not need hundreds of repeated simulations to obtain a good estimate of the early exercise boundary, but a much lower number, e.g., $I = 10$ suffices. Determining the early exercise boundary is the most computationally complex part of the algorithm since this involves performing several cross-sectional regressions, and our method therefore offers significant computational speedups. For example, one can run $I = 10$ initial simulations with $N = 10,000$ paths to estimate the optimal early exercise boundary. One can then use $N = 1,000,000$ paths for pricing, whereby an estimate with very low bias and low variance is obtained. Note that in this setup, obtaining the initial estimate of the optimal early exercise boundary takes no more time than running the standard LSM method once,

and since it is straightforward to parallelize our method, it could run in a fraction of this time on multicore processors or modern computer clusters.

It should be noted that several papers have proposed alternative refinements to this type of simulation method. In particular, it has been suggested to use well known variance reduction techniques like antithetic simulation, control variates, importance sampling, as well as initial dispersion together with the LSM method (see, among others: (Juneja and Kalra 2009; Lemieux and La 2005; Rasmussen 2005)). However, there are very few alternative suggestions for how to reduce the bias of the simulated estimates of American option prices in general and of the LSM method in particular. One potentially interesting method, which could be combined with our method, is that of Kan et al. (2009), although their application was to the value function iteration method of, e.g., Carriere (1996) and Tsitsiklis and Van Roy (2001). Our proposed method could also be used together with the inequality constrained least-squares method suggested in Létourneau and Stentoft (2014) or the method that corrects for heteroskedasticity in the cross-sectional regression proposed by Fabozzi et al. (2017).

Finally, our current applications are to standard financial options, but our results have important implications for other applications of simulation based pricing as well. One particularly important and challenging area in which the LSM method has been used is the field of real option valuation. For example, the work in Gamba and Fusari (2009) proposed a general valuation approach for capital budgeting decision involving modularization, and the work in Kang and Létourneau (2016) studied the investment decision and choice between coal and natural gas power plants under political risk with options to turn production on or off, while the work in Power et al. (2015) studied the valuation and timing of complex infrastructure projects. In all these cases, the regular LSM method was used to approximate complex decision problems in settings with multiple risk factors with complicated dynamics and for problems where the payoff function is non-standard and “exotic”. We conjecture that our proposed bootstrapping method will allow more efficient determination of the optimal controls and more precise valuation of these assets.

The remainder of the paper is structured as follows: Section 2 describes the discrete time framework used for valuation, introduces the least-squares Monte Carlo method of Longstaff and Schwartz (2001), and demonstrates how our proposed bootstrapping method can be implemented and improves estimation of the early exercise boundary. Section 3 presents an extensive numerical analysis of the performance of our proposed method demonstrating its robustness to changes in the simulation setup, to different option characteristics, and its applicability to higher dimensional problems. Section 4 provides a discussion of the potential benefits of separating the estimation of the early exercise boundary from the pricing and demonstrates that our method can be implemented in a very efficient way that works well in empirically relevant settings. Section 5 concludes.

2. Framework

The first step in implementing a numerical algorithm to price early exercise options is to assume that time can be discretized. We specify J exercise points as $t_0 = 0 < t_1 \leq t_2 \leq \dots \leq t_J = T$, with t_0 and T denoting the current time and maturity of the option, respectively. Thus, we are essentially approximating the American option by the so-called Bermudan option. The American option price is obtained in the limit by increasing the number of exercise points, J ; see also Bouchard and Warin (2012) for a formal justification of this approach.³ We assume a complete probability space $(\Omega, \mathcal{F}, \mathbb{Q})$ equipped with a discrete filtration $(\mathcal{F}(t_j))_{j=0}^J$ and a unique pricing measure corresponding to the probability measure \mathbb{Q} . The derivative’s value depends on one or more underlying assets modeled using a Markovian process, with state variables $(X(t_j))_{j=0}^J$ adapted to the filtration. We denote by $(Z(t_j))_{j=0}^J$ an adapted discounted payoff process for the derivative satisfying $Z(t_j) = \pi(X(t_j), t_j)$ for a suitable

³ We do not stress any further this difference as the literature on pricing early exercise options using simulation generally refers to these as American style options; see, e.g., Longstaff and Schwartz (2001).

function $\pi(\cdot, \cdot)$ assumed to be square integrable. This notation is sufficiently general to allow for non-constant interest rates through the appropriate definition of the state variables X and the payoff function π (see, e.g., [Glasserman 2004](#)). Following, e.g., [Karatzas \(1988\)](#) and [Duffie \(1996\)](#), in the absence of arbitrage, we can specify the American option price as:

$$P(X(0) = x) = \max_{\tau(t_1) \in \mathcal{T}(t_1)} E[Z(\tau) | X(0) = x], \tag{1}$$

where $\mathcal{T}(t_j)$ denotes the set of all stopping times with values in $\{t_j, \dots, t_J\}$. Thus, we explicitly assume that the option cannot be exercised at time $t = 0$.

The problem of calculating the option price in (1) with $J > 1$ is referred to as a discrete time optimal stopping time problem and typically solved using the dynamic programming principle. Intuitively this procedure can be motivated by considering the choice faced by the option holder at time t_j . The optimal choice will be to exercise immediately if the value of this is positive and larger than the expected payoff from holding the option until the next period and behaving optimally onwards. Let $V(X(t_j))$ denote the value of the option for state variables X at a time t_j prior to expiration and define $F(X(t_j)) \equiv E[Z(\tau(t_{j+1})) | X(t_j)]$ as the expected conditional payoff, where $\tau(t_{j+1})$ is the optimal stopping time. It then follows that:

$$V(X(t_j)) = \max(Z(t_j), F(X(t_j))), \tag{2}$$

and the optimal stopping time can be derived iteratively as:

$$\begin{cases} \tau(t_j) = T \\ \tau(t_j) = t_j \mathbf{1}_{\{Z(t_j) \geq F(X(t_j))\}} + \tau(t_{j+1}) \mathbf{1}_{\{Z(t_j) < F(X(t_j))\}}, \quad 1 < j \leq J - 1. \end{cases} \tag{3}$$

Based on this stopping time, the value of the option in (1) can be calculated as:

$$P(X(0) = x) = E[Z(\tau(t_1)) | X(0) = x]. \tag{4}$$

The backward induction theorem of [Chow et al. \(1971\)](#) (Theorem 3.2) provides the theoretical foundation for the algorithm in (3) and establishes the optimality of the derived stopping time and the resulting price estimate in (4).

2.1. Simulation and Regression Methods

The idea behind using simulation for option pricing is quite simple and involves estimating expected values and, therefore, option prices by an average of a number of random draws. However, when the option is American, one needs to determine simultaneously the optimal early exercise strategy, and this complicates matters. In particular, it is generally not possible to implement the exact algorithm in (3) because the conditional expectations are unknown, and therefore, the price estimate in (4) is infeasible. Instead, an approximate algorithm is needed. Because conditional expectations can be represented as a countable linear combination of basis functions, we may write $F(X(t_j)) = \sum_{m=0}^{\infty} \phi_m(X(t_j)) c_m(t_j)$, where $\{\phi_m(\cdot)\}_{m=0}^{\infty}$ form a basis.⁴ To make this operational we further assume that the conditional expectation function can be well approximated with the first

⁴ This is justified when approximating elements of the L^2 space of square integrable functions relative to some measure. Since L^2 is a Hilbert space, it has a countable orthonormal basis (see, e.g., [Royden 1988](#)).

$M + 1$ terms such that $F(X(t_j)) \approx F_M(X(t_j)) = \sum_{m=0}^M \phi_m(X(t_j)) c_m(t_j)$ and that we can obtain an estimate of this function by:

$$\hat{F}_M^N(X(t_j)) = \sum_{m=0}^M \phi_m(X(t_j)) \hat{c}_m^N(t_j), \tag{5}$$

where the coefficients $\hat{c}_m^N(t_j)$ are approximated or estimated using $N \geq M$ simulated paths. For example, in the Least-Squares Monte Carlo (LSM) method of Longstaff and Schwartz (2001), these are determined from a cross-sectional regression of the discounted future path-wise payoff on transformations of the state variables.

Based on the estimate in (5), we can derive an estimate of the optimal stopping time as:

$$\begin{cases} \hat{\tau}_M^N(t_J) = T \\ \hat{\tau}_M^N(t_j) = t_j \mathbf{1}_{\{Z(t_j) \geq \hat{F}_M^N(X(t_j))\}} + \hat{\tau}_M^N \mathbf{1}_{\{Z(t_j) < \hat{F}_M^N(X(t_j))\}}, \quad 1 < j \leq J - 1. \end{cases} \tag{6}$$

From the algorithm in (6), a natural estimate of the option value in (4) is given by:

$$\hat{P}_M^N(X(0) = x) = E[Z(\hat{\tau}_M^N(1)) | X(0) = x]. \tag{7}$$

In the special case when all the paths are started at the current values of the state variables, i.e., $X(0) = x$, the conditional expectation in (7) can be estimated by the sample average given by:

$$\hat{P}_M^N(X(0) = x) = \frac{1}{N} \sum_{n=1}^N Z(n, \hat{\tau}_M^N(1, n)), \tag{8}$$

where $Z(n, \hat{\tau}_M^N(1, n))$ is the payoff from exercising the option at the estimated optimal stopping time $\hat{\tau}_M^N(1, n)$ determined for path n according to (6). Convergence of this type of estimate has been analyzed in detail in the existing literature. The first step in doing so is to establish the convergence of the estimated approximate conditional expectation function, which is done in, e.g., the following lemma.

Lemma 1 (Adapted from Theorem 2 of Stentoft 2004b). *Under some regularity and integrability assumptions on the conditional expectation function, F (see Stentoft (2004b) for details), if $M = M(N)$ is increasing in N such that $M \rightarrow \infty$ and $M^3/N \rightarrow 0$, then $\hat{F}_M^N(X(t_j))$ converges to $F(X(t_j))$ in probability for $j = 1, \dots, J$.*

Proof. See Stentoft (2004b). \square

The result in Lemma 1 can now be combined with Proposition 1 of Stentoft (2004b) to demonstrate that when all the simulated paths are started at the current values of the state variable, i.e., $X(0) = x$, then the estimate in (8) converges to the true price, which establishes the convergence of the LSM method in a general multi-period setting.⁵ Moreover, this type of algorithm has nice properties, and the work in Stentoft (2014) documented that it is the most efficient method when compared to, e.g., the value function iteration method of Carriere (1996) or Tsitsiklis and Van Roy (2001). When paths are started at initially dispersed values, the LSM method still allows one to estimate the optimal early exercise boundary, and Lemma 1 continues to hold. However, when it comes to pricing the option

⁵ One of the important assumptions in Stentoft (2004b) is that the support is bounded. In Glasserman and Yu (2002), convergence was studied in the unbounded case. This complicates the analysis, and therefore, they limited their attention to the normal and lognormal cases. See also Gerhold (2011) for generalizations of the results to other processes and Belomestny (2011) for a proof using nonparametric local polynomial regressions.

with initially dispersed paths, this is more complicated, and an initial regression is needed at time $t = 0$; see, e.g., [Létourneau and Stentoft \(2019\)](#) for a proof that this method converges.

2.2. Regression and Optimal Early Exercise

It is explicit in Equation (3) that the optimal early exercise region is determined by the comparison of $Z(t_j)$ and $F(X(t_j))$ or in the case when the conditional expectations are approximated by the comparison of $Z(t_j)$ and $\hat{F}_M^N(X(t_j))$. In particular, we can define the true early exercise region implicitly as the value $B(t_j)$ that solves:

$$Z(t_j) > F(B(t_j)). \tag{9}$$

Similarly the estimated early exercise region is defined as the value $\hat{B}_M^N(t_j)$ that solves:

$$Z(t_j) > \hat{F}_M^N(\hat{B}_M^N(t_j)). \tag{10}$$

In the case where there is only one state variable, the exercise region at an exercise date is determined by a single point of intersection between the payoff function and conditional expectation. The work in [Rasmussen 2005](#) proposed a Newton–Raphson procedure to find the exercise boundary. In our application, we find this exercise boundary by subtracting the exercise value from the approximated conditional expectation and determining the roots of the resulting polynomial. Note that with an approximated conditional approximation, multiple roots may exist. If this happens, the largest root below the strike is kept as the exercise boundary.

While it is simple to find and represent the optimal early exercise boundary in the plain vanilla single option case, for more complex options, this might not be the case. For example, when there are two state variables, the conditional expectations and payoff functions are planes in a three-dimensional space. Hence, the intersection is a curve in a three-dimensional space, and it becomes much more difficult to represent the intersection.⁶ Moreover, while finding polynomial roots is easy in one dimension, this approach does not generalize easily to more complex situations. On the other hand, comparing the payoff function to the approximated conditional expectation is straightforward in multiple dimensions. Since the conditional expectations are completely described by the estimated coefficients $\hat{c}_m^N(t_j)$ from the cross-sectional regression, in practical implementations, it is therefore much easier to store these than it is to attempt to characterize and store a representation of the exercise region.

It follows from Lemma 1 above that as M and N tend to infinity, the estimated early exercise boundary will converge to the true early exercise boundary. For finite choices, however, $\hat{B}_M^N(t_j)$ is only an estimate of the true early exercise boundary, and the “quality” of this estimate will depend on the choice of M and N in a particular application. The fact that the early exercise boundary is estimated leads to two types of biases. First, when using a given estimated frontier, the option may be exercised at times when it should not have been or not exercised at times when it should have been. In both cases, this suboptimal early exercise will result in a low biased price estimate.⁷ Second, when using the same set of simulated paths to the approximation of the exercise boundary and to price the option, what we call in-sample pricing, the method suffers from a high bias from over-fitting the continuation value to the current sample of simulated paths. As explained first by [Broadie and Glasserman \(1997\)](#), the practice of using the same paths for the exercise decision and the payoff calculation introduces a positive correlation between the exercise decision and the future payoffs, essentially resulting in

⁶ For exotic options like stair-step options, there can be multiple intersections and multiple exercise regions.

⁷ Exercising when you should and not exercising when you should not have, on the other hand, no effect.

a foresight bias. The standard LSM estimator has both the low and the high bias, and therefore, the overall bias is difficult to sign.

2.3. Bootstrapping the Early Exercise Boundary

Although simulation and regression methods like the LSM method outlined above determine sequentially a number of conditional expectations as can be seen from Equation (6), these are used only to determine if the option at a given time should be exercised along a given path. As such, the conditional expectations are used as a convenient way to summarize, by subtracting the exercise value, the early exercise region and hence the path-wise optimal stopping time, which uniquely determines the option value. In Figure 1a, we plot in light gray $I = 100$ such estimated early exercise boundaries for a put option. As the figure shows, these are close to the true optimal early exercise boundary with the dotted black line although quite noisy, and this particularly so at the first steps in the simulation. Figure 1a however also shows that taking averages of the $I = 100$ conditional expectation approximations at any time and using that to back out the optimal early exercise boundary leads to a much smoother frontier shown with the dotted blue line and, supposedly, a better behaved estimated price.

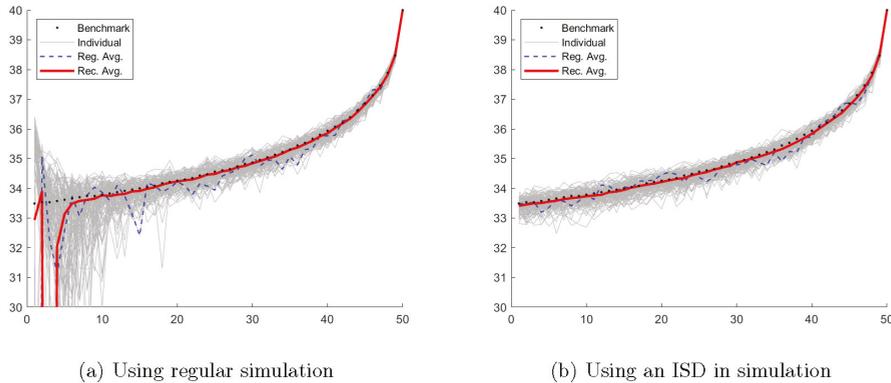


Figure 1. Individual, average, and bootstrapped early exercise boundaries. This figure shows $I = 100$ individually estimated early exercise boundaries using $N = 100,000$ paths of simulated data with $M = 9$ regressors and a constant term in the cross-sectional regressions. The right hand plot uses the proposed Initial State Dispersion (ISD) method from Rasmussen (2005). The option has a strike price of $K = 40$, a maturity of $T = 1$ year, and $J = 50$ early exercise points per year. The initial stock price is fixed at $S(0) = 40$; the volatility is $\sigma = 20\%$; and the interest rate is $r = 6\%$. The dotted blue line shows the early exercise boundary estimated from the average of the $I = 100$ continuation value approximations at each time and the red line the frontier backed out from our bootstrapped recursively averaged continuation values. The dashed black line shows the true early exercise boundary estimated with a binomial model with 50,000 steps.

Our proposed bootstrapping technique averages at each step in the backwards induction algorithm instead of only after having gone through all the steps in the backward induction algorithm. For example, at time $t = T - 1$, we estimate $I = 100$ independent conditional expectation approximations, i.e., polynomials of order $M = 9$ and a constant term in the cross-sectional regressions. Before proceeding backwards, we take the average of these approximations and use this same average across the $I = 100$ independent simulations to determine if the option along a given path in a given simulation at time $t = T - 1$ should be exercised or not. We then simply proceed, in a similar manner, backwards in time, always averaging at each early exercise point to time $t = 0$. In Figure 1a, we show in red the early exercise boundary backed out from these average conditional expectations at each

early exercise time. This estimated optimal early exercise boundary is for the most part extremely close to the true optimal early exercise boundary, and estimated option prices obtained with this estimate are therefore expected to be very close to the true option price.

In Table 1, we report the corresponding price estimates using both In-Sample (IS) pricing, i.e., when the same paths are used to determine the optimal early exercise boundary and for pricing, and Out-of-Sample (OS) pricing, where new simulated stock prices are used for pricing. A benefit of using OS pricing in the LSM method is that the price estimate is guaranteed to have a low expected bias because of the sub-optimality of the estimated optimal early exercise boundary. The first line in the table reports the relevant benchmark values for both the IS and OS method generated by applying the true early exercise boundary estimated from the binomial model with 50,000 steps to the relevant simulated paths. In this case, the IS and OS simply represent two different sets of paths. The table shows that as expected, the regular LSM method provides a high biased estimate when using IS and a low biased estimate when using OS, and these differences are in fact significant at standard levels.⁸ The difference between the two estimates is about half a cent, though slightly less when factoring in the Monte Carlo error as evidenced by the difference of 0.0013 between the benchmark price using the IS and OS sample. For the methods based on averages, the table shows that the estimated prices, both IS and OS, are much closer to the benchmark values and insignificantly different. This is particularly so for the recursive averages based on our bootstrapping method, which are spot on for the IS method and off by only a hundredth of a cent for the OS method. This indicates that for this setup, we have essentially obtained the true optimal early exercise strategy with our proposed bootstrapping method.

Table 1. Option prices using In-Sample (IS) and Out-of-Sample (OS) methods. LSM, Least-Squares Monte Carlo.

	Using IS Method		Using OS Method		Difference
Benchmark Boundary	2.3150	(0.0084)	2.3137	(0.0080)	0.0013
Individual LSM	2.3170	(0.0087)	2.3119	(0.0079)	0.0051
Regular Average	2.3148	(0.0085)	2.3136	(0.0079)	0.0013
Recursive Average	2.3150	(0.0084)	2.3138	(0.0080)	0.0012

This table shows the estimated prices and standard deviations using $I = 100$ individually estimated early exercise boundaries using $N = 100,000$ paths of simulated data with $M = 9$ regressors and a constant term in the cross-sectional regressions. The option has a strike price of $K = 40$, a maturity of $T = 1$ year, and $J = 50$ early exercise points per year. The initial stock price is fixed at $S(0) = 40$; the volatility is $\sigma = 20\%$; and the interest rate is $r = 6\%$. The binomial model estimate obtained with 50,000 steps is 2.3141. The benchmark boundary denotes the results from a method in which the true early exercise boundary estimated from the binomial model is used in the Monte Carlo simulation. Individual LSM denotes the regular method where the average of the $I = 100$ individual simulations are used. Regular average denotes the method where the average of the individually estimated optimal early exercise strategy is used, and recursive average denotes the method where the bootstrapped estimated optimal early exercise strategy is used. By comparing these results to the values from the benchmark boundary, the error coming from the Monte Carlo simulation is eliminated.

The very jagged paths early on in the simulations in Figure 1a are due in part to there being very few paths that are deep in the money, which makes it difficult to estimate the value for which exercise is optimal. This can be corrected or improved upon by using Initial State Dispersion (ISD) as suggested by, e.g., Rasmussen (2005). The resulting estimated early exercise boundaries are shown in Figure 1b. Compared to Figure 1a, we see that using an ISD helps quite a bit for the individual early exercise boundaries, which are quite a bit less volatile up to the 25th early exercise points. However, while using an ISD may help when estimating the individual optimal early exercise boundaries, comparing the two figures clearly shows that bootstrapping helps much more.

⁸ The one-sided p -value for a test of zero bias when compared to the estimate obtained with the benchmark boundary is 4.9% and 5.5% for the IS and OS prices, respectively.

The method we propose is simple to implement and yields very good results with finite choices of the number of regressors used in the simulation, the number of simulated paths, and the number of repeated independent simulations. A natural next step is to ask about the asymptotic properties of this algorithm. It is straightforward to show that our method provides asymptotically unbiased price estimates, which we state in the following corollary.

Corollary 1. *The bootstrapping method provides an asymptotically unbiased estimate of the option price for any choice of I , i.e., the number of repeated independent simulations used, under the assumptions outlined in Lemma 1.*

Proof. This follows by applying Lemma 1 to each of the independent simulations and noting that since the individual $\hat{F}_M^N(X(t_j))$ converge to $F(X(t_j))$ in probability for $j = 1, \dots, J$ so does the average of I such simulations. \square

3. Results

The previous section demonstrated how our bootstrapping method can be implemented and shows that it leads to much more precise estimates of the optimal early exercise boundary and results in estimated prices that are essentially unbiased for a benchmark option. In this section, we test the robustness of these findings along several dimensions. First, we show that our results are robust across choices in the simulation setup for the number of regressors and the number of paths used and across option characteristics like the moneyness and maturity of the option and the volatility of the underlying asset. To illustrate this, we consider first options on a single stock in a simple model with Black–Scholes–Merton dynamics because this allows us to characterize the true optimal early exercise boundary, which can be used to obtain Monte Carlo benchmark values. Finally, we show that our results generalize to the case with multiple underlying assets. We consider four different payoffs, and though the pricing performance varies across payoffs and depends on the order of the polynomial approximation, our bootstrapped method performs the best across different approximations.⁹

3.1. Robustness to the Simulation Setup

In Section 2.3, we demonstrated how to implement the bootstrapping method for a given number of regressors, $M = 9$, and number of simulated paths, $N = 100,000$. However, both M and N are choice parameters in the simulation setup that need to be picked when implementing the method. Thus, the first thing we analyze is the robustness of our proposed method to the choice of the number of simulated paths, N , used for determining the estimated optimal early exercise boundary and subsequently for pricing in case the OS method is used. The option we consider here has a strike price of $K = 40$, a maturity of $T = 1$ year, and $J = 50$ early exercise points per year. The initial stock price is fixed at $S(0) = 40$; the volatility is $\sigma = 20\%$; and the interest rate is $r = 6\%$. For now, we continue to use a polynomial of order $M = 9$ in the cross-sectional regressions.

Figure 2 compares the results from using the three different methods to determine the estimated optimal early exercise boundary: the regular LSM method, the regular average of the individual LSM boundaries, and the recursively bootstrapped boundaries, as a function of the number of simulated paths, N . The red lines represent the average price over $I = 100$ independent repetitions of the LSM method, each of which uses N paths. The line above the benchmark reference uses the IS method, whereas the line below uses the OS method. To eliminate the sampling bias from using a finite number of paths in the Monte Carlo simulation, we compare the price estimates to what would be obtained using the true early exercise boundary backed out from a binomial model with 50,000 steps on the two sets of simulated paths. The green lines use the regular average over the $I = 100$ individual LSM

⁹ All simulations were conducted using MATLAB.

repetitions using N paths, and the blue lines use the recursive average over $I = 100$ individual LSM repetitions using N paths.

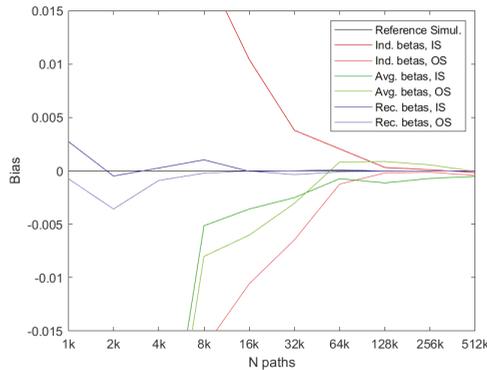


Figure 2. Convergence of price estimates as a function of sample size, N . This figure shows the price estimates from $I = 100$ simulations with different numbers of paths N . Individually, early exercise boundaries are estimated with $M = 9$ regressors and a constant term in the cross-sectional regressions. The option has a strike price of $K = 40$, a maturity of $T = 1$ year, and $J = 50$ early exercise points per year. The initial stock price is fixed at $S(0) = 40$; the volatility is $\sigma = 20\%$; and the interest rate is $r = 6\%$. The red lines report the results for the standard LSM method. The green lines report the results when the early exercise boundary is estimated from the average of the $I = 100$ continuation value approximations. The blue lines report the results when the early exercise boundary is backed out from our bootstrapped continuation values. The horizontal black line shows the result when the true early exercise boundary estimated with a binomial model with 50,000 steps is used.

First, the figure clearly shows that the regular IS LSM price can be strongly biased if the number of paths is too low. This is due to using the same paths to determine the individual optimal early exercise boundary and for calculating payoffs, which as we discussed earlier, leads to a high bias in the estimated option price. The OS prices, i.e., from the method that uses a new set of paths for pricing, are guaranteed to be low biased as shown in the figure. As the number of paths used in the individual simulation increases, the IS and OS results converge to a value that is (slightly) lower than the benchmark value, but corresponds to the true approximate value obtained when the approximation $F_M(X(t_j))$ based on a polynomial of order $M = 9$ of the true conditional expectation $F(X(t_j))$ is used.

Next, the figure shows that using the regular average method over the $I = 100$ repetitions improves on the results when a low number of paths is used. This is the situation where the early exercise boundary is estimated with the most noise, and averaging helps counter that somewhat. Note that the IS prices for this method are also low biased because of the averaging. As expected, both price estimates converge to the true approximated price as the number of paths used increases and in the limit, which in our setting corresponds to using $N = 512,000$ paths; averaging has very little effect on the estimated prices.

The final and most interesting thing the figure shows is that using the recursive average dramatically improves on the results across all values of N . In fact, Figure 2 essentially shows that the recursive method is virtually unbiased, even when using as little as $N = 1000$ paths to approximate the conditional expectations. Using only $N = 1000$ paths for OS pricing with the estimated early exercise boundary also leads to unbiased estimates, although both of these estimates will naturally have quite large variances.

In addition to N being a choice parameter for the LSM method, so is M . In Figure 3, we plot equivalent results to those in Figure 2, but using $M = 5$ and $M = 15$, respectively. The first thing to notice is that the two plots in Figure 3 show a pattern that is very similar to that obtained when using

$M = 9$ in Figure 2. For example, the LSM method is high biased when using IS pricing and low biased when using OS pricing with $M = 5$, as well as with $M = 15$. Figure 3, however, does indicate that these biases are larger for a given value of N when $M = 15$ than when $M = 5$. This is expected since the degree of overfitting increases with M . This also affects the method that averages the optimal early exercise boundaries somewhat, although our bootstrapping method is much less effected. In particular, our proposed method works very well for both $M = 5$ and $M = 15$, even when a low number of simulated paths, N , is used in the simulation.

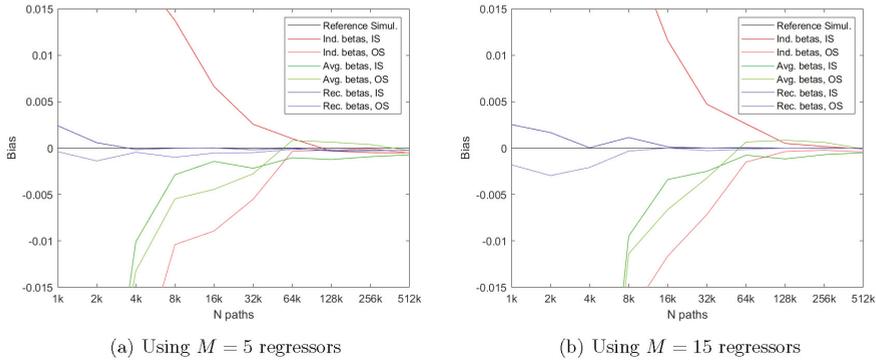


Figure 3. Convergence of price estimates for other values of regressors, M . This figure shows the price estimates from $l = 100$ simulations with different numbers of paths N . Individually, early exercise boundaries are estimated with $M = 5$ and $M = 15$, respectively, regressors,, and a constant term in the cross-sectional regressions. The option has a strike price of $K = 40$, a maturity of $T = 1$ year, and $J = 50$ early exercise points per year. The initial stock price is fixed at $S(0) = 40$; the volatility is $\sigma = 20\%$; and the interest rate is $r = 6\%$. The red lines report the results for the standard LSM method. The green lines report the results when the early exercise boundary is estimated from the average of the $l = 100$ continuation value approximations. The blue lines report the results when the early exercise boundary is backed out from our bootstrapped continuation values. The horizontal black line shows the result when the true early exercise boundary estimated with a binomial model with 50,000 steps is used.

3.2. Robustness across Option Characteristics

We now demonstrate that our previous findings are robust to considering options with different moneyness, maturity, and volatility of the underlying asset. To examine this, we consider additional options with strike prices of $K = 36$ and $K = 44$, maturities of $T = 0.5$ and $T = 2$ years, and volatilities of the underlying of $\sigma = 10\%$ and $\sigma = 40\%$. Again, we assume that the options have $J = 50$ early exercise points per year, such that the option with $T = 0.5$ years maturity has 25 early exercise points and the option with $T = 2$ years maturity has 100 early exercise points, respectively. We keep the number of regressors fixed at $M = 9$, but plot the resulting price estimates as a function of the number of simulated paths, N .

Figure 4 shows the performance of the three methods for an Out of The Money (OTM) option with $K = 36$ in the left hand plot and for an In The Money (ITM) option with $K = 44$ in the right hand plot. The first thing to notice is that the two plots in Figure 4 show a pattern that is very similar to that obtained for the At The Money (ATM) option with $K = 40$ in Figure 2. Figure 4 does indicate, though, that the biases in the LSM method are larger for a given value of N when pricing an OTM option than when pricing an ITM option. This is expected since the degree of overfitting increases as we go out of the money where less paths are used in the cross-sectional regressions. However, our proposed method continues to work very well, and much better than the regular LSM method, for both OTM and ITM options even when a low number of simulated paths, N , is used in the simulation.

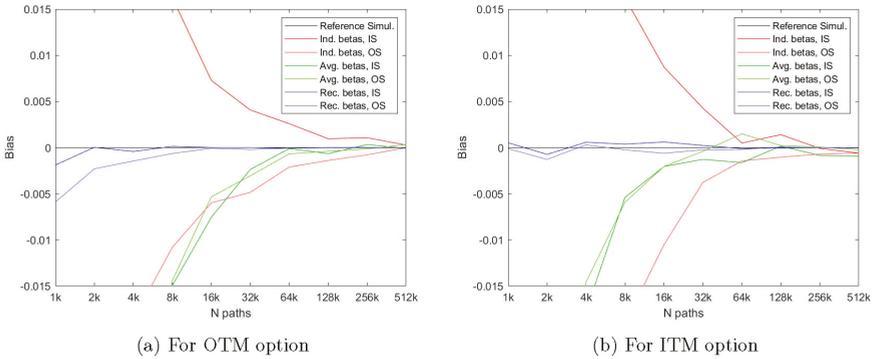


Figure 4. Convergence of price estimates for other values of the strike price, K . This figure shows the price estimates from $I = 100$ simulations with different numbers of paths N . Individually, early exercise boundaries are estimated with $M = 9$ regressors and a constant term in the cross-sectional regressions. The option has a strike price of $K = 36$ and $K = 44$, respectively, a maturity of $T = 1$ year, and $J = 50$ early exercise points per year. The initial stock price is fixed at $S(0) = 40$; the volatility is $\sigma = 20\%$; and the interest rate is $r = 6\%$. The red lines report the results for the standard LSM method. The green lines report the results when the early exercise boundary is estimated from the average of the $I = 100$ continuation value approximations. The blue lines report the results when the early exercise boundary is backed out from our bootstrapped continuation values. The horizontal black line shows the result when the true early exercise boundary estimated with a binomial model with 50,000 steps is used.

Figure 5 shows the performance of the three methods for a Short Maturity (ST) option with $T = 0.5$ years to maturity and 25 early exercise points in the left hand plot and for a Long Maturity (LT) option with $T = 2$ years to maturity and 100 early exercise points in the right hand plot. The first thing to notice is again that the two plots in Figure 5 show a pattern that is very similar to that obtained for the option with a maturity of $T = 1$ year in Figure 2. Figure 5 does indicate, though, that the biases in the LSM method are larger for a given value of N when the maturity of the option increases. This is to be expected since errors in the approximation of the conditional expectations accumulate in the backward algorithm, and we thus expect larger accumulated errors for longer maturities. However, our proposed method continues to work very well, and much better than the regular LSM method, for both ST and LT options even when a low number of simulated paths, N , is used in the simulation.

Figure 6 shows the performance of the three methods for an option on an asset with a low volatility of $\sigma = 10\%$ in the left hand plot and for an option on an asset with a high volatility of $\sigma = 40\%$ in the right hand plot. Again, the first thing to notice is that the two plots in Figure 6 show a pattern that is very similar to that obtained for the option on an underlying asset with a volatility of $\sigma = 20\%$ in Figure 2. Figure 6 does indicate, though, that the biases in the LSM method are larger for a given value of N when the volatility of the underlying asset increases. However, our proposed method continues to work very well, and much better than the regular LSM method, for both options on low and high volatility assets even when a low number of simulated paths, N , is used in the simulation.

3.3. Robustness to the Dimensionality of the Problem

We now demonstrate that our previous findings for options on a single asset are robust when increasing the number of underlying assets. To examine this, we consider options on three underlying assets with payoffs on the arithmetic average, the geometric average, the maximum, and the minimum of the underlying assets. We consider options that are at the money with $K = 40$ and a maturity of $T = 0.5$ years, and have $J = 25$ early exercises in total. The underlying asset prices are $S_i(0) = 40$; the volatilities are $\sigma_i = 40\%$, for $i = 1, 2, 3$; the correlation between all assets is $\rho = 0.5$; and the interest rate is $r = 6\%$. Benchmark prices are obtained with a binomial model with 2000 steps a year. Since it is

difficult to characterize explicitly the optimal early exercise boundary for these options, we compare to the true price, and we therefore cannot take the Monte Carlo error into consideration.

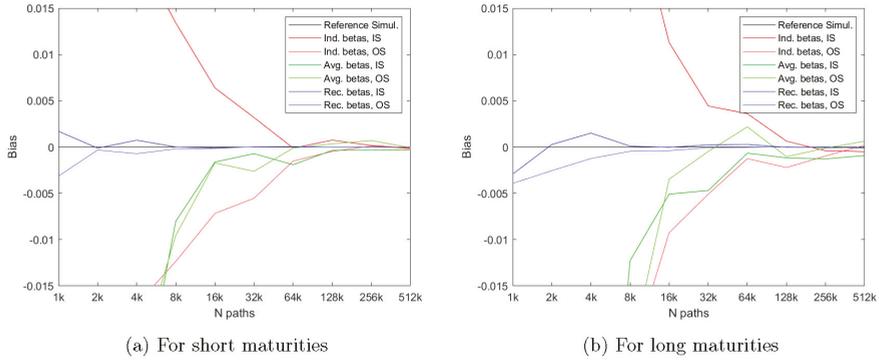


Figure 5. Convergence of price estimates for other maturities, T . This figure shows the price estimates from $I = 100$ simulations with different numbers of paths N . Individually, early exercise boundaries are estimated with $M = 5$ and $M = 15$ regressors and a constant term in the cross-sectional regressions, respectively. The option has a strike price of $K = 40$, a maturity of $T = 0.5$ and $T = 2$ years, respectively, and $J = 50$ early exercise points per year. The initial stock price is fixed at $S(0) = 40$; the volatility is $\sigma = 20\%$; and the interest rate is $r = 6\%$. The red lines report the results for the standard LSM method. The green lines report the results when the early exercise boundary is estimated from the average of the $I = 100$ continuation value approximations. The blue lines report the results when the early exercise boundary is backed out from our bootstrapped continuation values. The horizontal black line shows the result when the true early exercise boundary estimated with a binomial model with 50,000 steps is used.

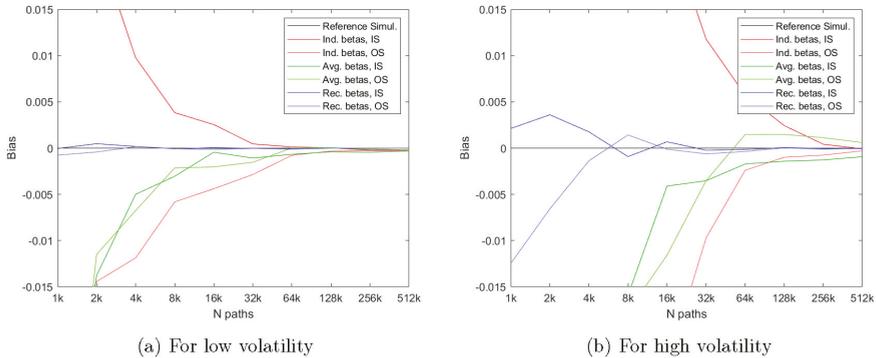


Figure 6. Convergence of price estimates for other values of the volatility, σ . This figure shows the price estimates from $I = 100$ simulations with different numbers of paths N . Individually early exercise boundaries are estimated with $M = 9$ regressors and a constant term in the cross-sectional regressions. The option has a strike price of $K = 40$; a maturity of $T = 1$ year; and $J = 50$ early exercise points per year. The initial stock price is fixed at $S(0) = 40$, the volatility is $\sigma = 10\%$ and $\sigma = 40\%$, respectively, and the interest rate is $r = 6\%$. The red lines report the results for the standard LSM method. The green lines report the results when the early exercise boundary is estimated from the average of the $I = 100$ continuation value approximations. The blue lines report the results when the early exercise boundary is backed out from our bootstrapped continuation values. The horizontal black line shows the result when the true early exercise boundary estimated with a binomial model with 50,000 steps is used.

Figure 7 shows the price estimates for options with the four different payoff across the number of paths used in the simulation, N . In all cases, the complete set of polynomials of order $M = 9$ and a

constant term are used as regressors in the cross-sectional regressions. The first thing to note is that in all cases the algorithms converge, though for options on the maximum and minimum, this is to a somewhat low biased estimate. This makes sense since the conditional expectations are more difficult to approximate for these options; see also [Stentoft \(2004a\)](#). More importantly, though, for all the payoffs, our proposed bootstrapping method delivers the least biased price estimates. Thus, the results clearly show that our proposed method is robust to increases in the dimension of the pricing problem.

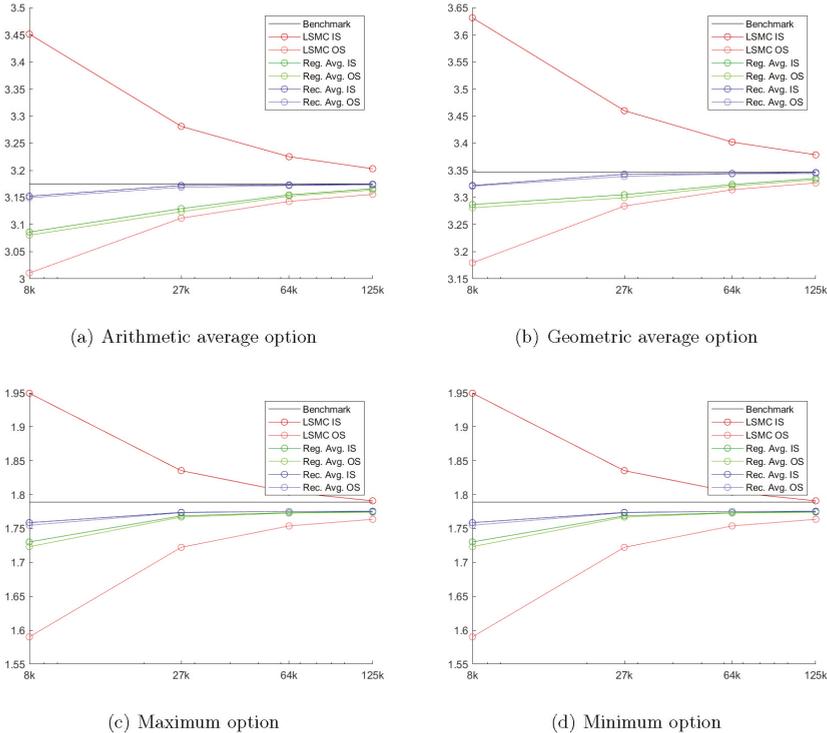


Figure 7. Convergence of price estimates for options on three assets. This figure shows the price estimates from $I = 100$ simulations with different numbers of paths N . Individually, early exercise boundaries are estimated with the complete set of polynomials of order $M = 9$ and a constant term as regressors in the cross-sectional regressions. The option has a strike price of $K = 40$, a maturity of $T = 0.5$ year, and $J = 50$ early exercise points per year. The initial stock price is fixed at $S(0) = 40$; the volatility is $\sigma = 40\%$; and the interest rate is $r = 6\%$. The red lines report the results for the standard LSM method. The green lines report the results when the early exercise boundary is estimated from the average of the $I = 100$ continuation value approximations. The blue lines report the results when the early exercise boundary is backed out from our bootstrapped continuation values. The horizontal black line shows the benchmark price obtained with a binomial model with 2000 steps per year.

To demonstrate further our method’s robustness, [Figure 8](#) shows the corresponding price estimates when the maximum order of the complete polynomial is $M = 15$, in which case a total of 816 regressors are used in the cross-sectional regressions. It is possible that by using other regressors, e.g., functions of the maximum asset value, than the complete set of polynomials, better results could be obtained with less regressors; however, for consistency and simplicity, we chose to stay with monomials as the basis. The figure shows that when M is increased, the asymptotic bias, apparent when using a very large number of simulated paths N , decreases, and this is particularly so for the maximum and minimum

options. Note though that for all payoffs, our proposed bootstrapping method again continues to perform the best and deliver price estimates with the lowest bias of all the methods reported.

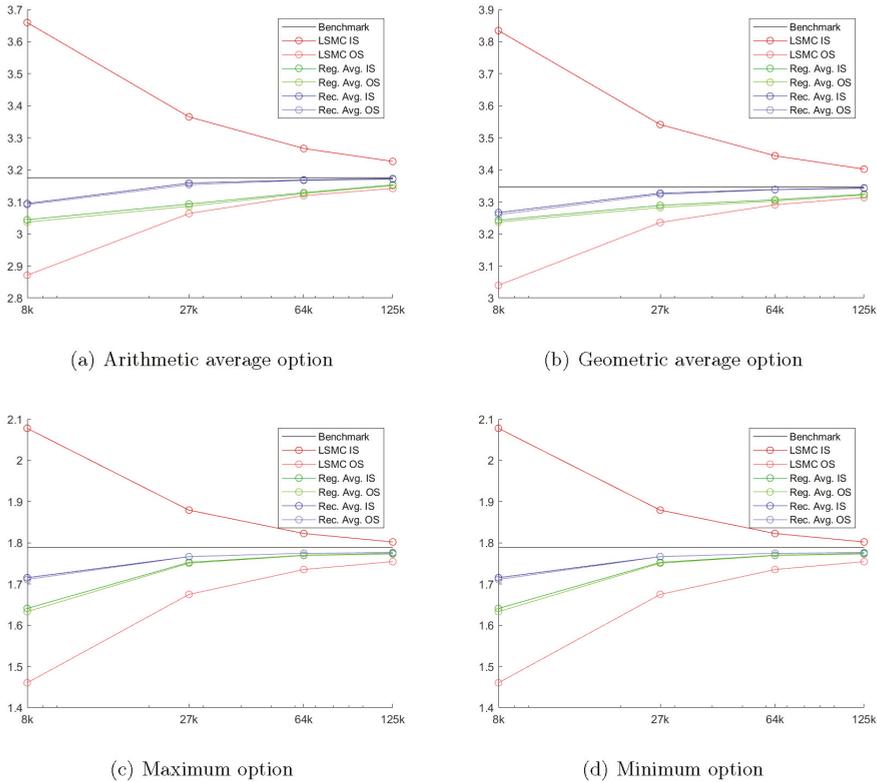


Figure 8. Convergence of price estimates for options on three assets, $M = 15$. This figure shows the price estimates from $I = 100$ simulations with different numbers of paths N . Individually, early exercise boundaries are estimated with the complete set of polynomials of order $M = 15$ and a constant term as regressors in the cross-sectional regressions. The option has a strike price of $K = 40$, a maturity of $T = 0.5$ year, and $J = 50$ early exercise points per year. The initial stock price is fixed at $S(0) = 40$; the volatility is $\sigma = 40\%$; and the interest rate is $r = 6\%$. The red lines report the results for the standard LSM method. The green lines report the results when the early exercise boundary is estimated from the average of the $I = 100$ continuation value approximations. The blue lines report the results when the early exercise boundary is backed out from our bootstrapped continuation values. The horizontal black line shows the benchmark price obtained with a binomial model with 2000 steps per year.

4. Discussion

The results in Section 3 were obtained by averaging over $I = 100$ repeated samples with the same number of simulated paths, N , for both IS and OS pricing. In this section, we first demonstrate that it is in fact not necessary to average over such a large number of repeated samples when bootstrapping the optimal early exercise boundary. We also show that it is in fact not necessary to average over independently repeated simulations with a large number of simulated paths in the bootstrapping

method either.¹⁰ Finally, since the number of repeats, I_{IS} , and the number of simulated paths, N_{IS} , used for IS pricing can be disassociated from the number of repeats, I_{OS} , and number of simulated paths, N_{OS} , used for OS pricing, we propose to price options using our proposed bootstrapping method with reasonably low values for I_{IS} and N_{IS} , since this is enough to obtain unbiased results, and large values of, in particular, N_{OS} , since this will deliver price estimates with a low variance.¹¹

Figure 9 shows that the quality of the approximation of the estimated early exercise boundary very quickly improves, reflected in decreasing absolute bias of the estimated price, as I_{IS} and N_{IS} increase. In fact, if as little as $I_{IS} = 10$ repeats and $N_{IS} = 10,000$ simulated paths are used, the OS bias is very small, and when $I_{IS} = 10$ repeats and $N_{IS} = 50,000$ simulated paths are used, it is essentially eliminated. Note that estimating the optimal early exercise boundary with $I_{IS} = 10$ repeats and $N_{IS} = 10,000$ simulated paths can be done in roughly the same time as running the regular LSM method once with $N = 100,000$, a number typically used.¹²

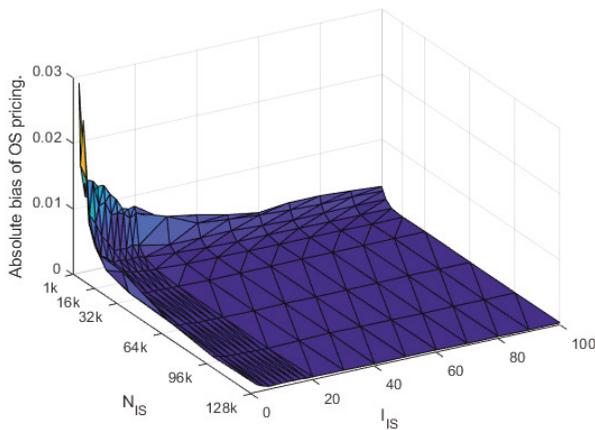


Figure 9. OS bias in the bootstrapping method This figure shows the out-of-sample bias obtained using $I_{OS} = 100$ and $N_{OS} = 100,000$ paths when using early exercise boundaries determined with different numbers of in sample repeats, I_{IS} , and paths, N_{IS} , using our proposed bootstrapping method with a polynomial of order $M = 9$ and a constant term as regressors in the cross-sectional regressions. The option has a strike price of $K = 40$, a maturity of $T = 1$ year, and $J = 50$ early exercise points per year. The initial stock price is fixed at $S(0) = 40$; the volatility is $\sigma = 20\%$; and the interest rate is $r = 6\%$.

To illustrate the efficiency of our method, we now report for all combinations of moneyness, maturity, and volatility, a total of 27 options, the estimated prices from implementing our bootstrapping method with $I_{IS} \ll I_{OS}$ and $N_{IS} < N_{OS}$. Specifically, we set $I_{IS} = 10$ and $N_{IS} = 50,000$ and price the options out of sample with $I_{OS} = 100$ and $N_{OS} = 100,000$, the standard choice in the literature. Compared to the regular LSM method, the IS results can be calculated in a fraction of the time (5% roughly) with our bootstrapping method. The results are reported in Table 2, which compares the bootstrapped results to the results that would be obtained had the true optimal early exercise boundaries been used.

¹⁰ Given the results in, e.g., Figure 2, this should come as no surprise.

¹¹ Note that since the same approximations are used for all OS simulations, it does not matter if we simulate $I = 100$ times with 100,000 paths or once with $N = 10,000,000$ paths.

¹² Note that our proposed bootstrapping method is straightforward to implement on multiple cores, and more generally, it can be implemented on clusters to take advantage of available parallel computing resources.

Table 2. Option prices using bootstrapping.

K	T	σ	Benchmark Boundary		Recursive Average		Difference
			Price	St. Dev.	Price	St. Dev.	
36	0.5	10%	0.0304	(0.0006)	0.0303	(0.0006)	−0.0000
36	1	10%	0.0896	(0.0012)	0.0894	(0.0012)	−0.0001
36	2	10%	0.1715	(0.0018)	0.1714	(0.0018)	−0.0001
36	0.5	20%	0.4974	(0.0035)	0.4971	(0.0035)	−0.0003
36	1	20%	0.9169	(0.0052)	0.9166	(0.0052)	−0.0003
36	2	20%	1.4317	(0.0069)	1.4311	(0.0071)	−0.0006
36	0.5	40%	2.1978	(0.0104)	2.1972	(0.0102)	−0.0006
36	1	40%	3.4360	(0.0135)	3.4355	(0.0134)	−0.0006
36	2	40%	4.9619	(0.0168)	4.9606	(0.0169)	−0.0013
40	0.5	10%	0.7344	(0.0029)	0.7343	(0.0029)	−0.0000
40	1	10%	0.8892	(0.0031)	0.8891	(0.0032)	−0.0001
40	2	10%	1.0235	(0.0038)	1.0234	(0.0038)	−0.0001
40	0.5	20%	1.7907	(0.0072)	1.7907	(0.0071)	−0.0001
40	1	20%	2.3137	(0.0080)	2.3133	(0.0079)	−0.0005
40	2	20%	2.8833	(0.0099)	2.8824	(0.0099)	−0.0009
40	0.5	40%	3.9708	(0.0144)	3.9697	(0.0144)	−0.0010
40	1	40%	5.3115	(0.0170)	5.3103	(0.0168)	−0.0011
40	2	40%	6.9149	(0.0210)	6.9131	(0.0210)	−0.0017
44	0.5	10%	3.9474	(0.0019)	3.9474	(0.0019)	−0.0000
44	1	10%	3.9473	(0.0018)	3.9472	(0.0018)	−0.0001
44	2	10%	3.9480	(0.0020)	3.9480	(0.0019)	−0.0000
44	0.5	20%	4.3079	(0.0094)	4.3078	(0.0093)	−0.0001
44	1	20%	4.6528	(0.0098)	4.6523	(0.0098)	−0.0005
44	2	20%	5.0804	(0.0112)	5.0796	(0.0116)	−0.0008
44	0.5	40%	6.3232	(0.0176)	6.3228	(0.0176)	−0.0005
44	1	40%	7.6107	(0.0200)	7.6095	(0.0197)	−0.0011
44	2	40%	9.1784	(0.0232)	9.1764	(0.0233)	−0.0020

This table shows estimated prices and standard deviations using our proposed bootstrapping method. The OS method is used with $I_{OS} = 100$ and $N_{OS} = 100,000$ with 15 recursive averages calculated with $I_{JS} = 10$ and $N_{JS} = 50,000$ with $M = 9$ regressors, and a constant term in the cross-sectional regressions. The option characteristics are given in the first 3 columns. In all cases, the initial stock price is fixed at $S(0) = 40$; the options have $J = 50$ early exercise points per year; and the interest rate is $r = 6\%$. The benchmark boundary denotes the results from a method in which the true early exercise boundary estimated from the binomial model with 50,000 steps is used in the Monte Carlo simulation. By comparing the results to the values from the benchmark boundary, the error coming from the Monte Carlo simulation is eliminated. The differences are shown in the last column.

Table 2 shows that all the estimated prices are low biased, which is expected since we are using OS pricing. However, the absolute size of the bias is indeed very small, and in all cases, the bias is statistically insignificant. Across the 27 options, the largest bias in absolute value is -0.0020 , well below a cent, and the average bias across the sample of options is -0.0005 . Moreover, the table shows that the standard deviations of our estimated prices are similar to what is obtained when applying the benchmark frontier. Thus, when pricing this sample of diverse and empirically relevant options, our bootstrapping method essentially yields unbiased price estimates that are as precise as if the true optimal early exercise boundary had been used. If the regular LSM method had been used instead, the corresponding biases, both the largest ones and the average across options, would have been much larger and more volatile.¹³ Note that the option for which the bias is the largest is the long term in the money option on a high volatility underlying asset. The results in Section 3 demonstrate that this is the most challenging option to price with the regular LSM method, and in light of this, our bootstrapping method performs remarkably well.

The constant volatility Gaussian models considered above may not be adequate for empirical option pricing. Thus, to check the robustness of our results to more realistic alternatives, we now consider models with time varying volatility of the GARCH type. The work in Duan (1995) was

¹³ These results are available from the authors upon request.

among the first to show how to price options in a (Gaussian) GARCH model, and this framework has since been widely used empirically. See [Christoffersen et al. \(2013\)](#) for a detailed survey of the use of GARCH option pricing models. In the GARCH option pricing model, returns under the pricing measure \mathbb{Q} are given by:

$$R_t = r - \frac{1}{2}h_t + \varepsilon_t, \tag{11}$$

where $\varepsilon_t \mid \mathcal{F}_{t-1} \sim N(0, h_t)$, with \mathcal{F}_t denoting the information set at time t and where the conditional variance, h_t , follows a NGARCH process given by:

$$h_{t+1} = \omega + \beta h_t + \alpha \left(\varepsilon_t - (\lambda + \gamma) \sqrt{h_t} \right), \tag{12}$$

where ω, β, α , and γ are parameters governing the dynamics under the physical measure \mathbb{P} and λ is the constant unit risk premium. The GARCH model is obtained when $\gamma = 0$, and the constant volatility model amounts to setting all parameters except ω equal to zero.

Table 3 shows the results for three different volatility specifications and thus examines the robustness of our results to using more general stochastic processes. The first thing to notice from the table is that our proposed bootstrapping method generates price estimates with insignificant biases irrespective of the dynamic model used. Thus, the table demonstrates that the excellent performance of our proposed method does not depend on the complexity of the dynamics. Using the ordinary LSM method, however, leads to significantly biased price estimates, and this is particularly so when the conditional volatility follows the more complicated NGARCH process. Our proposed method, on the other hand, continues to deliver statistically insignificant price estimates when more complicated and empirically relevant dynamics are used, and the estimates are much less biased than with the ordinary LSM method.

Table 3. Option prices with time varying volatility.

Panel A: Constant Volatility Model								
Individual LSM					Recursive Average			
K	Price	St. Dev.	Bias	T-stat	Price	St. Dev.	Bias	T-stat
36	0.4983	(0.0036)	-0.0011	-2.8281	0.4990	(0.0036)	-0.0004	-1.0196
40	1.7936	(0.0068)	-0.0021	-2.9162	1.7945	(0.0066)	-0.0012	-1.6947
44	4.3158	(0.0075)	-0.0018	-2.2127	4.3171	(0.0073)	-0.0005	-0.6237
Panel B: GARCH Model								
Individual LSM					Recursive Average			
K	Price	St. Dev.	Bias	T-stat	Price	St. Dev.	Bias	T-stat
36	0.5060	(0.0037)	-0.0021	-5.4783	0.5076	(0.0037)	-0.0005	-1.3202
40	1.7725	(0.0068)	-0.0031	-4.3681	1.7747	(0.0069)	-0.0009	-1.2259
44	4.2844	(0.0074)	-0.0030	-3.7376	4.2873	(0.0074)	-0.0001	-0.1330
Panel C: NGARCH model								
Individual LSM					Recursive Average			
K	Price	St. Dev.	Bias	T-stat	Price	St. Dev.	Bias	T-stat
36	0.5822	(0.0045)	-0.0026	-5.4823	0.5842	(0.0044)	-0.0005	-1.1326
40	1.7431	(0.0072)	-0.0035	-4.6303	1.7454	(0.0071)	-0.0012	-1.5857
44	4.1799	(0.0072)	-0.0036	-4.7237	4.1826	(0.0072)	-0.0010	-1.2543

This table shows estimated prices and standard deviations using the individual LSM method and our proposed bootstrapping method for models with time varying volatility of the GARCH type. The OS method is used with $I_{OS} = 100$ and $N_{OS} = 100,000$ with the complete set of polynomials of order $M_S = 5$ and $M_V = 3$ in the stock price and the volatility, respectively, in the cross-sectional regressions. The individual LSM uses the same numbers in the IS method, whereas the recursive averages are obtained with $I_{IS} = 10$ and $N_{IS} = 50,000$. The option strike prices are given in the first column, and all options have a maturity of $T = 0.5$ year and $J = 252$ early exercise points per year. The initial stock price is fixed at $S(0) = 40$; the volatility is $\sigma = 20\%$; and the interest rate is $r = 6\%$. The following parameters were used: $\beta = 0.92$, $\alpha = 0.05$, $\gamma = 0.5$, and $\lambda = 0.1$, and to ensure a risk neutral unconditional annual volatility of 20%, we set $\omega = \left(1 - \beta - \alpha \left(1 + (\lambda + \gamma)^2\right)\right) * 0.20 / \sqrt{252}$. Benchmark prices are calculated using $N = 1,000,000$ paths and the complete set of polynomials of order $M_S = 9$ and $M_V = 5$ in the stock price and the volatility, respectively, in the cross-sectional regressions.

5. Conclusions

This paper proposed an innovative algorithm that significantly improves on the approximation of the optimal early exercise boundary obtained with simulation based methods for American option pricing like, e.g., the Least-Squares Monte Carlo (LSM) method of Longstaff and Schwartz (2001). The method successfully exploited and leveraged the information in multiple cross-sectional regressions to the fullest by averaging the individually obtained estimates at each early exercise step, starting from just before maturity, in the backwards induction algorithm. We referred to our method as a bootstrapping approach because of the similarities it had with how the term structure of interest rates was bootstrapped and the way bootstrapping was used for inference in regression models. With the bootstrapping approach, less errors were accumulated in the backward induction algorithm, and as a result of this, the price estimate was essentially unbiased even for long maturity options.

We compared the results from our bootstrapping approach to the regular LSM method, and the numerical results demonstrated large and significant improvements from our method. These findings were robust to the choice of simulation setup, the characteristics of the option, and the dimensionality of the problem. Finally, because our method naturally disassociated the estimation of the optimal early exercise boundary from the pricing of the option, significant efficiency gains could be obtained by using less simulated paths and repetitions to estimate the optimal early exercise boundary than with the regular method. To illustrate this, we priced a diverse sample of options with different moneyness, maturity, and levels of volatility of the underlying asset. The results showed that when pricing this sample of empirically relevant options, our bootstrapping method essentially yielded unbiased price estimates that were as precise as if the true optimal early exercise boundary had been used.

Our bootstrapping method should have wide applications empirically. First, the majority of the options traded are in fact American style, and the recent global financial crisis clearly demonstrated that considering more risk factors is essential to model the complex behavior of financial markets properly. Our proposed method can be used with such advanced models and will allow more efficient pricing than what is currently possible. Second, our results have important implications for other simulation based applications. One particularly important and challenging area in which the LSM method has been used is the field of real option valuation. In particular, the dynamics of the underlying assets for this type of options are often very complicated, and simulation is the only viable method. We conjecture that our proposed bootstrapping method will allow more efficient determination of the optimal controls and more precise valuation of these assets, benefiting the economy at large.

Besides the empirical applications outlined above, there are other important lines for future research. First, we are currently working on establishing the finite sample properties of our proposed bootstrapping method to demonstrate theoretically its usefulness and to assess if there are any limitations to the applicability of the methodology. Second, simulation based methods have recently been used to obtain estimates of option risk sensitivities or hedge ratios, the Greeks, for American options. We conjecture that our bootstrapping method would allow such quantities to be estimated with much more precision as well and are currently working on establishing this empirically. Finally, our general idea could be applied to other methods that use simulation and regression techniques for pricing early exercise style option like, e.g., the value function iteration method of Carriere (1996) and Tsitsiklis and Van Roy (2001), and to other dynamic programming methods used to solve optimal control type problems.

Author Contributions: The authors contributed equally to all aspects of this paper.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank two anonymous referees for valuable comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Belomestny, Denis. 2011. Pricing Bermudan Options by Nonparametric Regression: Optimal Rates of Convergence for Lower Estimates. *Finance and Stochastics* 15: 655–83. [CrossRef]
- Bouchard, Bruno, and Xavier Warin. 2012. Monte-Carlo Valuation of American Options: Facts and new Algorithms to Improve Existing Methods. In *Numerical Methods in Finance*. Bordeaux: Springer, pp. 215–55.
- Boyer, Martin M., and Lars Stentoft. 2013. If We Can Simulate It, We Can Insure It: An Application to Longevity Risk Management. *Insurance: Mathematics and Economics* 52: 35–45. [CrossRef]
- Broadie, Mark, and Paul Glasserman. 1997. Pricing American-style Securities using Simulation. *Journal of Economic Dynamics and Control* 21: 1323–52. [CrossRef]
- Carriere, Jacques F. 1996. Valuation of the Early-Exercise Price for Options using Simulations and Nonparametric Regression. *Insurance: Mathematics and Economics* 19: 19–30. [CrossRef]
- Chow, Yuan-Shih, Herbert Robbins, and David Siegmund. 1971. *Great Expectations: The Theory of Optimal Stopping*. New York: Houghton Mifflin.
- Christoffersen, Peter, Kris Jacobs, and Chayawat Ornthanalai. 2013. GARCH Option Valuation: Theory and Evidence. *The Journal of Derivatives* 21: 8–41. [CrossRef]
- Duan, Jin-Chuan 1995. The GARCH Option Pricing Model. *Mathematical Finance* 5: 13–32.
- Duffie, Darrell 1996. *Dynamic Asset Pricing Theory*. Princeton: Princeton University.
- Fabozzi, Frank J., Tommaso Paletta, and Radu Tunaru. 2017. An Improved Least Squares Monte Carlo Valuation Method Based on Heteroscedasticity. *European Journal of Operational Research* 263: 698–706. [CrossRef]
- Gamba, Andrea, and Nicola Fusari. 2009. Valuing Modularity as a Real Option. *Management Science* 55: 1877–96. [CrossRef]
- Gerhold, Stefan 2011. The Longstaff–Schwartz Algorithm for Lévy Models: Results on Fast and Slow Convergence. *Annals of Applied Probability* 21: 589–608.
- Glasserman, Paul 2004. *Monte Carlo Methods in Financial Engineering*. New York: Springer.
- Glasserman, Paul, and Bin Yu. 2002. Simulation for American Options: Regression Now or Regression Later? In *Monte Carlo and Quasi-Monte Carlo Methods 2002*. Edited by Harald Niederreiter. Berlin: Springer.
- Ibanez, Alfredo, and Fernando Zapatero. 2004. Monte Carlo Valuation of American Options through Computation of the Optimal Exercise Frontier. *Journal of Financial and Quantitative Analysis* 39: 253–75. [CrossRef]
- Juneja, Sandeep, and Himanshu Kalra. 2009. Variance Reduction Techniques for Pricing American Options Using Function Approximations. *Journal of Computational Finance* 12: 79–102. [CrossRef]
- Kan, K. H. Felix, R. Mark Reesor, Tyson Whitehead, and Matt Davison. 2009. Correcting the Bias in Monte Carlo Estimators of American-style Option Values. In *Monte Carlo and Quasi-Monte Carlo Methods 2008*. Edited by Pierre L'Ecuyer and Art B. Owen, Berlin: Springer, pp. 439–54.
- Kang, Sang Baum, and Pascal Létourneau. 2016. Investors Reaction to the Government Credibility Problem: A Real Option Analysis of Emission Permit Policy Risk. *Energy Economics* 54: 96–107. [CrossRef]
- Karatzas, Ioannis. 1988. On the Pricing of American Options. *Applied Mathematics and Optimization* 17: 37–60. [CrossRef]
- Lemieux, Christiane, and Jennie La. 2005. A Study of Variance Reduction Techniques for American Option Pricing. Paper presented at the 2005 Winter Simulation Conference, Orlando, FL, USA, December 4–7, pp. 1884–91.
- Létourneau, Pascal, and Lars Stentoft. 2014. Refining the Least Squares Monte Carlo Method by Imposing Structure. *Quantitative Finance* 14: 495–507. [CrossRef]
- Létourneau, Pascal, and Lars Stentoft. 2019. *Simulated Greeks for American Options, Working Paper*. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3503889 (accessed on 15 November 2019).
- Longstaff, Francis A., and Eduardo S. Schwartz. 2001. Valuing American Options by Simulation: A Simple Least-Squares Approach. *Review of Financial Studies* 14: 113–47. [CrossRef]
- Moreno, Manuel, and Javier F. Navas. 2003. On the Robustness of Least-Squares Monte Carlo (LSM) for Pricing American Derivatives. *Review of Derivatives Research* 6: 107–28. [CrossRef]
- Power, Gabriel J., Charli D. Tandja M, Josée Bastien, and Philippe Grégoire. 2015. Measuring Infrastructure Investment Option Value. *The Journal of Risk Finance* 16: 49–72. [CrossRef]
- Rasmussen, Nicki. 2005. Control Variates for Monte Carlo Valuation of American Options. *Journal of Computational Finance* 9: 2–12. [CrossRef]
- Royden, Halsey. 1988. *Real Analysis*. Upper Saddle River: Prentice Hall, Inc.

- Stentoft, Lars. 2004a. Assessing the Least Squares Monte-Carlo Approach to American Option Valuation. *Review of Derivatives Research* 7: 129–68. [[CrossRef](#)]
- Stentoft, Lars. 2004b. Convergence of the Least Squares Monte Carlo Approach to American Option Valuation. *Management Science* 9: 1193–203. [[CrossRef](#)]
- Stentoft, Lars. 2014. Value Function Approximation or Stopping Time Approximation: A Comparison of Two Recent Numerical Methods for American Option Pricing using Simulation and Regression. *Journal of Computational Finance* 18: 1–56. [[CrossRef](#)]
- Tsitsiklis, John N., and Benjamin Van Roy. 2001. Regression Methods for Pricing Complex American-Style Options. *IEEE Transactions on Neural Networks* 12: 694–703. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Article

Generalized Mean-Reverting 4/2 Factor Model

Yuyang Cheng [†], Marcos Escobar-Anel ^{*†} and Zhenxian Gong [†]

Department of Statistical and Actuarial Sciences, Western University, London, ON N6A 3K7, Canada; yche259@uwo.ca (Y.C.); zgong26@uwo.ca (Z.G.)

* Correspondence: marcos.escobar@uwo.ca

[†] These authors contributed equally to this work.

Received: 13 September 2019; Accepted: 26 September 2019; Published: 8 October 2019

Abstract: This paper proposes and investigates a multivariate 4/2 Factor Model. The name 4/2 comes from the superposition of a CIR term and a 3/2-model component. Our model goes multidimensional along the lines of a principal component and factor covariance decomposition. We find conditions for well-defined changes of measure and we also find two key characteristic functions in closed-form, which help with pricing and risk measure calculations. In a numerical example, we demonstrate the significant impact of the newly added 3/2 component (parameter b) and the common factor (a), both with respect to changes on the implied volatility surface (up to 100%) and on two risk measures: value at risk and expected shortfall where an increase of up to 29% was detected.

Keywords: stochastic covariance; 4/2 model; option pricing; risk measures

1. Introduction

Continuous-time stochastic covariance models are crucial in capturing many stylized facts in financial data, from heteroscedasticity and fat tails to changing correlations and leverage effects. Early work in this field focused on discrete time models in the form of generalized autoregressive conditional heteroskedasticity (GARCH) models (see Engle 2002). The best-known representatives in continuous time, are the stochastic Wishart family (see Da Fonseca et al. 2007; Gouriéroux 2006) and the Ornstein–Uhlenbeck (OU) family (see Muhle-Karbe et al. 2012) of models, as well as general linear-quadratic jump-diffusions (see Cheng and Scaillet 2007). These approaches are more realistic than the classical Black–Scholes lognormal model, but they quickly become intractable as dimensions increase in terms of the number of parameters and simulation paths, commonly known as the “curse of dimensionality”. Recent papers (see De Col et al. 2013; Escobar 2018) have presented models built from linear combination of tractable one-dimensional counterparts. These models involve fewer parameters than Wishart- or OU-type approaches, owing to a reduction in dimensionality while providing a closed-form solution to financial problems.

In this paper, we introduce a multivariate mean-reverting stochastic volatility factor model that combines 1/2 (Heston-type, Heston 1993) and 3/2 processes (Platen 1997) for the modeling of volatility. Such underlying volatility processes were coined 4/2 by Grasselli (2017). Our paper takes advantage of the factor structure in asset prices and allows for a mean-reverting structure on the assets thereby aiming at capturing either multivariate commodity behavior or multiple volatility indexes (see Gnoatto et al. 2018 for an alternative multivariate non-mean-reverting generalization based on a pairwise-structure applied to the exchange-rate market). In particular, our setting reduces the dimension of the parametric space which is a way of controlling the “curse of dimensionality” making parameters identifiable and popular estimation methods feasible. Secondly, the presence of independent common and intrinsic factors, each with its own stochastic volatility, enables an elegant separable structure for characteristic functions (c.f.s) and captures several stylized facts, such as: stochastic volatility, stochastic correlation among stocks (see Engle 2002), co-movements in the

variances (see [Diebold and Nerlove 1989](#)), multiple factors in the volatilities (see [Heston et al. 2009](#)) and stock correlations (see [Da Fonseca et al. 2007](#)). Thirdly, a factor representation is compatible with economical interpretations, where common factors are exogenous variables explaining financial markets, and intrinsic factors relate to companies’ intrinsic risks. Lastly, closed-form expressions are available for joint c.f.s; this is useful for derivative pricing and risk management calculations via Fourier transformations, and it makes c.f.-based estimations methods feasible (see [Carr and Madan 1999](#); [Caldana and Fusai 2013](#); [Fusai et al. 2018](#)).

The rationale for a 4/2 volatility process rather than a 1/2 or 3/2 model is masterly presented in [Grasselli \(2017\)](#) for a one-dimensional structure. For instance, as observed by the author, the 1/2 process predicts that the implied volatility skew will flatten when the instantaneous volatility increases (crises), while the 3/2 model predicts steepening skews. The empirical violation of the Feller condition in the 1/2 model is also noted, which makes volatility paths stay closer to 0 for a longer period than empirically supported, while the 3/2 model admits extreme paths with spikes in instantaneous volatility. The two processes complement each other as they imply very different dynamics for the evolution of the implied volatility surface. It stands to reason that such a convenient underlying drive for multidimensional structures should be used to improve not only marginal volatility behavior, but also the dependence structure.

We obtain an analytical representation for the c.f. of the vector of asset prices, which is in closed-form for non-mean-reverting nested cases. This type of c.f. is helpful for derivative pricing purposes. We also produce a second conditional c.f. that can be used for exact simulations of the non-mean reverting assets given the terminal volatilities, where the latter can be simulated exactly via chi-squares. We identify a set of conditions that not only produces well-defined changes of measure, but also avoids local martingales; hence, it can be used for risk-neutral pricing purposes.

Our results were applied numerically to parameters inspired by commodity prices¹. There is a vast literature on commodity modeling (see, for instance, [Chiarella et al. 2013](#); [Schwartz 1997](#), and more recently [Schneider and Tavin 2018](#)). In our numerical study, we investigated the impact of the new parameters (b , the weight of 3/2 in the overall instantaneous volatility) on the shape of the implied volatility surface and the values of two risk measures: VaR and expected shortfall.

2. Model Description

Next, we define the model in a filtered probability space $(\Omega, \mathcal{F}, \mathcal{P}, \mathbb{F})$ where \mathcal{F}_0 contains all subsets of the $(\mathcal{P}-)$ null sets of \mathcal{F} and \mathbb{F} is right-continuous. We first provide the processes under the historical measure \mathcal{P} , then followed by the processes under a (conveniently chosen) risk-neutral measure \mathcal{Q} . Suppose that $X(t) = (X_1(t), \dots, X_n(t))'$ is a vector of asset prices with the following \mathcal{P} -measure representation:

$$\begin{aligned} \frac{dX_i(t)}{X_i(t)} &= \left\{ L_i + c_i \sum_{j=1}^p a_{ij}^2 \left(\sqrt{v_j(t)} + \frac{b_j}{\sqrt{v_j(t)}} \right)^2 - \sum_{j=1}^n \beta_{ij} \ln(X_j(t)) + \tilde{c}_i \left(\sqrt{\tilde{v}_i(t)} + \frac{\tilde{b}_i}{\sqrt{\tilde{v}_i(t)}} \right)^2 \right\} dt \\ &\quad + \sum_{j=1}^p a_{ij} \left(\sqrt{v_j(t)} + \frac{b_j}{\sqrt{v_j(t)}} \right) dW_j(t) + \left(\sqrt{\tilde{v}_i(t)} + \frac{\tilde{b}_i}{\sqrt{\tilde{v}_i(t)}} \right) d\tilde{W}_i(t) \\ dv_j(t) &= \alpha_j(\theta_j - v_j(t))dt + \zeta_j \sqrt{v_j(t)} dB_j^P(t), \quad j = 1, \dots, p \\ d\tilde{v}_i(t) &= \tilde{\alpha}_i(\tilde{\theta}_i - \tilde{v}_i(t))dt + \tilde{\zeta}_i \sqrt{\tilde{v}_i(t)} d\tilde{B}_i^P(t), \quad i = 1, \dots, n \end{aligned}$$

¹ This can also be applied to volatility indexes, such as those reported by the Chicago Board Options Exchange (CBOE), which are clearly a mean-reverting asset class with stochastic volatility.

The quadratic variation structure is $\langle dB_j^P(t), dW_j(t) \rangle = \rho_j dt$, $\langle d\tilde{B}_i^P(t), d\tilde{W}_i(t) \rangle = \tilde{\rho}_i dt$ and zero otherwise. In the language of factor analysis, a_{ij} is the ij th entry of the matrix of factor loadings (A) that captures the correlations among assets. The communalities are represented by $V_j(t) = \left(\sqrt{v_j(t)} + \frac{b_j}{\sqrt{v_j(t)}} \right)^2$ (in matrix form, $\Lambda_{n \times p} = A \text{diag}(V^{1/2}(t))$) and the intrinsic residual variance is $\Psi = \text{diag}(\tilde{V}(t))$, with $\tilde{V}_i(t) = \left(\sqrt{\tilde{v}_i(t)} + \frac{\tilde{b}_i}{\sqrt{\tilde{v}_i(t)}} \right)^2$. This leads to a factors decomposition of the quadratic variation of asset prices as follows:

$$\Sigma(t)dt = (\Lambda\Lambda' + \Psi) dt = \left(A \text{diag}(V(t))A' + \text{diag}(\tilde{V}(t)) \right) dt$$

Whenever necessary, we assume $n = p$ and $A = (a_{ij})_{n \times p}$ to be an orthogonal matrix. In this setting, c_i and \tilde{c}_i represent risk premiums of asset $X_i(t)$ associated with the common and intrinsic factors, respectively. $\beta = (\beta_{ij})_{n \times n}$ is an invertible matrix, which captures the spillover at the expected return level $X_i(t)$ on asset $X_j(t)$. In other words, it represents the impact from other assets on the long term average price of the current one.

Based on the quadratic variation relationship defined in this model, if we assume that $B_j^P, B_j^P(t)^\perp, \tilde{B}_i^P(t), \tilde{B}_i^P(t)^\perp$ are independent Brownian motions with $-1 \leq \rho_j \leq 1$ and $-1 \leq \tilde{\rho}_i \leq 1$. Then,

$$dW_j(t) = \rho_j dB_j^P(t) + \sqrt{1 - \rho_j^2} dB_j^P(t)^\perp$$

$$d\tilde{W}_i(t) = \tilde{\rho}_i d\tilde{B}_i^P(t) + \sqrt{1 - \tilde{\rho}_i^2} d\tilde{B}_i^P(t)^\perp.$$

$v_j, j = 1, \dots, n$ and $\tilde{v}_i, i = 1, \dots, n$ follow standard CIR processes, hence α_j, θ_j , and ξ_j are positive constants satisfying $\alpha_j \theta_j \geq \frac{\xi_j^2}{2}$ (the Feller condition). Similarly, $\tilde{\alpha}_i, \tilde{\theta}_i$, and $\tilde{\xi}_i$ are positive constants satisfying $\tilde{\alpha}_i \tilde{\theta}_i \geq \frac{\tilde{\xi}_i^2}{2}$. Note that the Feller condition in CIR model guarantees that the process remains positive.

The transformation $Y = \ln X$ would create a multivariate Ornstein–Uhlenbeck process with a 4/2 stochastic factor structure:

$$dY_i(t) = \left\{ L_i + \left(c_i - \frac{1}{2} \right) \sum_{j=1}^n a_{ij}^2 \left(\sqrt{v_j(t)} + \frac{b_j}{\sqrt{v_j(t)}} \right)^2 - \sum_{j=1}^n \beta_{ij} Y_j(t) + \left(\tilde{c}_i - \frac{1}{2} \right) \left(\sqrt{\tilde{v}_i(t)} + \frac{\tilde{b}_i}{\sqrt{\tilde{v}_i(t)}} \right)^2 \right\} dt + \sum_{j=1}^n a_{ij} \left(\sqrt{v_j(t)} + \frac{b_j}{\sqrt{v_j(t)}} \right) dW_j(t) + \left(\sqrt{\tilde{v}_i(t)} + \frac{\tilde{b}_i}{\sqrt{\tilde{v}_i(t)}} \right) d\tilde{W}_i(t)$$

This general model includes a notable particular case, which is a direct generalization of [Grasselli \(2017\)](#) to a factor setting when $\beta_{ij} = 0, i, j = 1, \dots, n$; this case is studied in more detail and it is named **FG**, given its analytical flexibility.

3. Results

This section describes viable changes of measure and two key characteristic functions of the targeted multivariate process; one for pricing and the other for simulations. The proofs are presented in [Appendix A](#).

3.1. Change of Measure

Here, we explore the topic of creating a risk-neutral measure \mathcal{Q} for pricing purposes. As noted by [Grasselli \(2017\)](#); [Platen and Heath \(2010\)](#) and [Baldeaux et al. \(2015\)](#) among others, a risk-neutral measure may not be supported by data in the presence of a 3/2 model (e.g., $\frac{1}{\sqrt{v(t)}}$), as the parametric

constraints needed for the discounted asset price process to be a \mathcal{Q} -martingale are violated with real data; hence, we can only produce a strict \mathcal{Q} -local martingale (i.e., \mathcal{Q} would be absolute continuous but not equivalent to \mathcal{P}). In such situation, the standard risk-neutral pricing methodology would fail (biased prices), and we have to turn to the benchmark approach for pricing (see Baldeaux et al. 2015).

The next proposition entertains the following changes of measure with constant $\lambda_j, \lambda_j^\perp, \tilde{\lambda}_i$ and $\tilde{\lambda}_i^\perp$ (see Escobar and Gong 2019 for other types of changes of measures) then identifies the parametric conditions needed for the existence of a valid risk-neutral measure \mathcal{Q} .

$$dB_j^{\mathcal{Q}}(t) = \lambda_j \left(\sqrt{v_j(t)} + \frac{b_j}{\sqrt{v_j(t)}} \right) dt + dB_j^{\mathcal{P}}(t), d\tilde{B}_i^{\mathcal{Q}}(t) = \tilde{\lambda}_i \left(\sqrt{\tilde{v}_i(t)} + \frac{\tilde{b}_i}{\sqrt{\tilde{v}_i(t)}} \right) dt + d\tilde{B}_i^{\mathcal{P}}(t)$$

$$dB_j^{\mathcal{Q}}(t)^\perp = \lambda_j^\perp \left(\sqrt{v_j(t)} + \frac{b_j}{\sqrt{v_j(t)}} \right) dt + dB_j^{\mathcal{P}}(t)^\perp, d\tilde{B}_i^{\mathcal{Q}}(t)^\perp = \tilde{\lambda}_i^\perp \left(\sqrt{\tilde{v}_i(t)} + \frac{\tilde{b}_i}{\sqrt{\tilde{v}_i(t)}} \right) dt + d\tilde{B}_i^{\mathcal{P}}(t)^\perp$$

Proposition 1. *The change of measure is well-defined for pricing purposes under the following four conditions:*

$$\xi_j^2 \leq 2\alpha_j\theta_j - 2\xi_j \max \left\{ |b_j\lambda_j|, |\lambda_j^\perp b_j|, |b_j a_{1j}\rho_j|, \dots, |b_j a_{nj}\rho_j| \right\} \tag{1}$$

$$\tilde{\xi}_i^2 \leq 2\tilde{\alpha}_i\tilde{\theta}_i - 2\tilde{\xi}_i \max \left\{ |\tilde{b}_i\tilde{\lambda}_i|, |\tilde{\lambda}_i^\perp\tilde{b}_i|, |\tilde{b}_i\tilde{\rho}_i| \right\} \tag{2}$$

$$\max \left\{ |\lambda_j|, |\lambda_j^\perp| \right\} < \frac{\alpha_j}{\xi_j} \tag{3}$$

$$\max \left\{ |\tilde{\lambda}_i|, |\tilde{\lambda}_i^\perp| \right\} < \frac{\tilde{\alpha}_i}{\tilde{\xi}_i} \tag{4}$$

Moreover, if $\beta_{ij} = 0$ for $i, j = 1, \dots, n$, then the following must also be satisfied:

$$L_i = r, c_i = \sum_{j=1}^n a_{ij} \left(\rho_j \lambda_j + \sqrt{1 - \rho_j^2} \lambda_j^\perp \right), \tilde{c}_i = \tilde{\rho}_i \tilde{\lambda}_i + \sqrt{1 - \tilde{\rho}_i^2} \tilde{\lambda}_i^\perp \tag{5}$$

Proof is included in Appendix A.

3.2. Characteristic Function

This section aims at obtaining an analytical representation for the c.f. If $Z(t) = e^{\beta t} Y(t)$ is defined such that $e^{\beta t}$ is a matrix exponential, then $Z_i(t)$ is represented as:

$$dZ_i(t) = \sum_{j=1}^n (e^{\beta t})_{ij} \left\{ L_i + \left(c_j - \frac{1}{2} \right) \sum_{k=1}^n a_{jk}^2 \left(\sqrt{v_k(t)} + \frac{b_k}{\sqrt{v_k(t)}} \right)^2 + \left(\tilde{c}_j - \frac{1}{2} \right) \left(\sqrt{\tilde{v}_j(t)} + \frac{\tilde{b}_j}{\sqrt{\tilde{v}_j(t)}} \right)^2 \right\} dt \tag{6}$$

$$+ \sum_{j=1}^n (e^{\beta t})_{ij} \left\{ \sum_{k=1}^n a_{jk} \left(\sqrt{v_k(t)} + \frac{b_k}{\sqrt{v_k(t)}} \right) dW_k(t) + \left(\sqrt{\tilde{v}_j(t)} + \frac{\tilde{b}_j}{\sqrt{\tilde{v}_j(t)}} \right) d\tilde{W}_j(t) \right\}$$

For convenience, we use $(e^{\beta t})_{ij}$ as the ij component of the matrix $e^{\beta t}$. Note that $Z_i(t)$ is no longer a mean-reverting process although it accounts for time dependent coefficients.

Next, we find the conditional c.f. for the increments of Z , defined as

$$\Phi_{Z(t),v(t)}(T, \omega) = E \left[\exp \left\{ i\omega' (Z(T) - Z(t)) \right\} \mid Z(t) = z_t, v(t) = v_t \right] \tag{7}$$

Under a risk neutral measure, this c.f. can be used for pricing some financial products, given the integrability conditions (a discussion of the generalized c.f. as per Grasselli 2017 is beyond the scope of this paper.). For convenience, we formulate it as $v(t) = (v_1(t), \dots, v_n(t), \tilde{v}_1(t), \dots, \tilde{v}_n(t))$.

Proposition 2. Let $(Z(t))_{t \geq 0}$ evolve according to the model in Equation (6). The c.f. $\Phi_{Z(t),v(t)}$ is then given as follows:

$$\begin{aligned} \Phi_{Z(t),v(t)}(T, \omega) &= E [\exp i\omega'(Z(T) - Z(t)) \mid Z(t) = y_t, v(t) = v_t] \\ &= \prod_{k=1}^n \Phi_{GG} \left(T, 1; L_k(\omega), h_k(\omega), g_k(\omega), \kappa_k, \theta_k, \xi_k, \rho_k, b_k, c_k, v_{k,t}, S_{k,t}^* \right) \\ &\quad \times \prod_{j=1}^n \Phi_{GG} \left(T, 1; 0, L_j(\omega), h_j(\omega), g_j(\omega), \tilde{\kappa}_j, \tilde{\theta}_j, \tilde{\xi}_j, \tilde{\rho}_j, \tilde{b}_j, \tilde{c}_k, \tilde{v}_{j,t}, S_{j,t}^{*j} \right) \end{aligned}$$

where Φ_{GG} is a one-dimensional generalization of the c.f. from Grasselli (2017) provided in Lemma A1.

Proof is provided in Appendix A. The c.f. above involves single expected values with respect to Brownian motion $B(t)$. In each term, Φ_{GG} (i.e., the second set of Brownian $W(t)$) is eliminated, hence this is a drastic simplification compared to the original $2n$ dimensional joint expectation.

A particular, fully solvable case is the FG model ($\beta_{ij} = 0, i, j = 1, \dots, n$).

Corollary 1. Let $(Z(t))_{t \geq 0}$ evolve according to the FG model ($\beta_{ij} = 0, i, j = 1, \dots, n$). The c.f. $\Phi_{Z(t),v(t)}$ is subsequently presented as follows:

$$\begin{aligned} \Phi_{Z(t),v(t)}(T, \omega) &= E [\exp i\omega'(Z(T) - Z(t)) \mid Z(t) = y_t, v(t) = v_t] \\ &= \prod_{k=1}^n \Phi_G \left(T, 1; L_k(\omega), h_k(\omega), g_k(\omega), \kappa_k, \theta_k, \xi_k, \rho_k, b_k, c_k, v_{k,t}, S_{k,t}^* \right) \\ &\quad \times \prod_{j=1}^n \Phi_G \left(T, 1; 0, L_j(\omega), h_j(\omega), g_j(\omega), \tilde{\kappa}_j, \tilde{\theta}_j, \tilde{\xi}_j, \tilde{\rho}_j, \tilde{b}_j, \tilde{c}_k, \tilde{v}_{j,t}, S_{j,t}^{*j} \right) \end{aligned}$$

where Φ_G is the one-dimensional c.f. provided by Grasselli (2017) in Proposition 3.1 and given in the Appendix B for completeness.

See Appendix A for proof. Next, we turn to the conditional c.f. of the increments of Z given the terminal value of the CIR processes. This is defined as follows:

$$\Phi_{Z(t),v(T)}(\tau, \omega) = E [\exp [\omega'(Z(T) - Z(t))] \mid Z(t) = z_t, v(T) = v_T] \tag{8}$$

The above is useful when we need to work with the joint distribution of $(Z(T), v(T))$ given $(Z(t), v(t))$. For such cases, we can try to rely on a convenient simulation scheme combining the distribution of $Z(T)$ given $(Z(t), v(T))$ (via Equation (8)) with that of $v(T)$ given $v(t)$, the latter is known to be non-centered chi-squared. In this way, we can avoid usual discretization algorithms such as the Euler–Maruyama or Milstein schemes, which are generally not suitable for the CIR process (due to failure of the Lipschitz condition at 0).

In this vein, when working with the non mean-reverting factor model ($\beta_{ij} = 0, i, j = 1, \dots, n$), we can easily adapt the procedures in Grasselli (2017) to provide an exact simulation scheme for the model given the vector of the independent CIR process at maturity T (i.e., $v(T)$). This requires only the c.f. provided next:

Corollary 2. Let $(Z(t))_{t \geq 0}$ evolve according to the FG model ($\beta_{ij} = 0, i, j = 1, \dots, n$). Then, the c.f. $\Phi_{Z(t),v(T)}$ is then given as follows:

$$\begin{aligned} \Phi_{Z(t),v(T)}(T, \omega) &= \prod_{j=1}^n \Phi_{G,T} \left(T, \phi; L, h_j, g_j, \kappa_j, \theta_j, \xi_j, \rho_j, b_j, c_j, v_{j,T}, S_{j,t}^* \right) \\ &\quad \times \prod_{i=1}^n \Phi_{G,T} \left(T, 1; 0, h_i, g_i, \tilde{\kappa}_i, \tilde{\theta}_i, \tilde{\xi}_i, \tilde{\rho}_i, \tilde{b}_i, \tilde{c}_i, \tilde{v}_{i,T}, S_t^{*i} \right) \end{aligned}$$

where $\Phi_{G,T}$ is the one-dimensional c.f. provided by Grasselli (2017) in Proposition 4.1 and given in the Appendix B for completeness.

Proof of this result is provided in Appendix A. Unsurprisingly, the previous result cannot be extended to the mean-reverting case, due to the absence of closed formulas for the object:

$$\mathbb{E} \left[\exp \left\{ u \left(\int_t^T B(s)v(s)ds + \int_t^T C(s) \frac{1}{v(s)} ds + \int_t^T D(s) \ln(v(s))ds \right) \right\} \mid v(T) \right]$$

which is not solvable even when two of the three deterministic functions $B(s)$, $C(s)$ and $D(s)$ are zero.

4. Discussion: One Common Factor in Two Dimensions

We assume two assets, i.e., $X_1(t)$ and $X_2(t)$, with one common volatility component, and one intrinsic factor each. The asset prices thereby follow the system of SDE for $i = 1, 2$:

$$\begin{aligned} dY_i(t) &= (L_i - \beta_i Y_i(t)) dt \\ &\quad + \left((c_i - \frac{1}{2}) [a_i^2 (\sqrt{v_1(t)} + \frac{b_1}{\sqrt{v_1(t)}})^2] + (\tilde{c}_i - \frac{1}{2}) (\sqrt{\tilde{v}_i(t)} + \frac{\tilde{b}_i}{\sqrt{\tilde{v}_i(t)}})^2 \right) dt \\ &\quad + a_i \left(\sqrt{v_1(t)} + \frac{b_1}{\sqrt{v_1(t)}} \right) dW_1(t) + \left(\sqrt{\tilde{v}_i(t)} + \frac{\tilde{b}_i}{\sqrt{\tilde{v}_i(t)}} \right) d\tilde{W}_i(t) \end{aligned}$$

$$\begin{aligned} dv_1(t) &= \alpha_1 (\theta_1 - v_1(t)) dt + \xi_1 \sqrt{v_1(t)} dB_1(t) \\ d\tilde{v}_i(t) &= \tilde{\alpha}_i (\tilde{\theta}_i - \tilde{v}_i(t)) dt + \tilde{\xi}_i \sqrt{\tilde{v}_i(t)} d\tilde{B}_i(t) \end{aligned}$$

with $\langle dB_j(t), dW_j(t) \rangle = \rho_j dt, \langle d\tilde{B}_i(t), d\tilde{W}_i(t) \rangle = \tilde{\rho}_i dt$ for $j = 1; i = 1, 2$.

The following table (Table 1) gives a baseline parameter set for the one-factor, two-dimensional 4/2 factor model used in the subsequent sections. The choice of parameters in Scenario A was made by combining the seminal works of Schwartz (1997) (see Oil and Copper in Tables IV and V) and Heston (1993). Scenario B combines Schwartz (1997) (see Oil and Copper, Tables IV and V) with Heston et al. (2009). In both cases, we assume a simple structure for the market price of risk ($c_1 = c_2 = \tilde{c}_1 = \tilde{c}_2 = 0$)².

² Variations on c will be studied in future research as part of a calibration exercise (see Medvedev and Scaillet 2007 for viable approaches and Escobar and Gschnaidtner 2016 for some pitfalls).

The $\tilde{\theta}_i$, $i = 1, 2$ in the table are set to match the long term volatilities as estimated in Schwartz (1997), which are 0.334 (Oil, Table IV) and 0.233 (Copper, Table V):

$$\mathbb{E} \left[a_1^2 \left(\sqrt{v_1(t)} + \frac{b_1}{\sqrt{v_1(t)}} \right)^2 + \left(\sqrt{\tilde{v}_1(t)} + \frac{\tilde{b}_1}{\sqrt{\tilde{v}_1(t)}} \right)^2 \right] = a_1^2 \left(\frac{2\alpha_1 b_1^2}{2\alpha_1 \theta_1 - \xi_1^2} + 2b_1 + \theta_1 \right) + \frac{2\tilde{\alpha}_1 \tilde{b}_1^2}{2\tilde{\alpha}_1 \tilde{\theta}_1 - \tilde{\xi}_1^2} + 2\tilde{b}_1 + \tilde{\theta}_1 = (0.334)^2 \tag{9}$$

This explains the values of $\tilde{\theta}_i$ in the table.

Table 1. Toy parametric values.

Initial Values
$X_1(0) = 18, X_2(0) = 100$
$v_1(0) = \theta_1, \tilde{v}_1(0) = \tilde{\theta}_1, \tilde{v}_2(0) = \tilde{\theta}_2$
Commodity Drift, Schwartz (1997)
$\beta_{11} = 0.301, \beta_{12} = 0, \beta_{21} = 0, \beta_{22} = 0.369$
$L_1 = 3.09\beta_{11} = 0.93, L_2 = 4.85\beta_{22} = 1.79$
Commodity St. Volatility, Heston (1993); Schwartz (1997). Scenario A
$\alpha_1 = \tilde{\alpha}_1 = \tilde{\alpha}_2 = 2$
$\theta_1 = 0.01, \tilde{\theta}_1 = 0.0753, \tilde{\theta}_2 = 0.0124$
$\xi_1 = \tilde{\xi}_1 = \tilde{\xi}_2 = 0.1$
$\rho_1 = \tilde{\rho}_1 = \tilde{\rho}_2 = -0.5$
Commodity St. Volatility, Heston et al. (2009); Schwartz (1997). Scenario B
$\alpha_1 = \tilde{\alpha}_1 = \tilde{\alpha}_2 = 0.2098$
$\theta_1 = 0.1633, \tilde{\theta}_1 = 0.0685, \tilde{\theta}_2 = 0.0689$
$\xi_1 = \tilde{\xi}_1 = \tilde{\xi}_2 = 0.1706$
$\rho_1 = \tilde{\rho}_1 = \tilde{\rho}_2 = -0.9$
New parameters.
$c_1 = c_2 = \tilde{c}_1 = \tilde{c}_2 = 0$
$a_1 = a_2 = 0.75$
$b_1 = \tilde{b}_1 = \tilde{b}_2 = 0.008$

The present section considers two independent cases. First, we study the impact of the parameters b_1, \tilde{b}_1 and \tilde{b}_2 on implied volatility surfaces and on two risk measures for a portfolio of underlyings. We then assess the impact of the commonalities a_1 and a_2 on these same targets, i.e., implied volatilities and risk measures. To ensure that the cases lead to reasonable assets behavior, we report the expected return, variance of return for each asset, as well as the correlation between two assets and the leverage effects in Tables 2 and 3 under Scenarios A and B, respectively.

We simulated 500,000 paths with $dt = 0.1$ and considered the following scenarios for b : $b_1 = 0.008, \tilde{b}_1 = \tilde{b}_2 = 0; b_1 = 0, \tilde{b}_1 = \tilde{b}_2 = 0.008; b_1 = \tilde{b}_1 = \tilde{b}_2 = 0$ and $b_1 = \tilde{b}_1 = \tilde{b}_2 = 0.008$.

Table 2. First four moments for scenarios on 3/2 component (b), Scenario A.

	$b_1 = 0.008, \tilde{b}_1 = \tilde{b}_2 = 0$	$b_1 = 0, \tilde{b}_1 = \tilde{b}_2 = 0.008$	$b_1 = \tilde{b}_1 = \tilde{b}_2 = 0$	$b_1 = \tilde{b}_1 = \tilde{b}_2 = 0.008$
$\mathbb{E} \left[\frac{X_1(T) - X_1(0)}{X_1(0)} \right]$	0.0494	0.0489	0.0503	0.0480
$\mathbb{E} \left[\frac{X_2(T) - X_2(0)}{X_2(0)} \right]$	0.0760	0.0754	0.0764	0.0750
$\mathbb{V} \left[\frac{X_1(T) - X_1(0)}{X_1(0)} \right]$	0.0663	0.0680	0.0618	0.0729
$\mathbb{V} \left[\frac{X_2(T) - X_2(0)}{X_2(0)} \right]$	0.0367	0.0390	0.0323	0.0445
$\text{Corr}(\ln X_1(T), \ln X_2(T))$	0.3194	0.0896	0.1060	0.2799
$\text{Corr}(\ln X_1(T), < \ln X_1(T) >)$	-0.4287	-0.4520	-0.4443	-0.4406
$\text{Corr}(\ln X_2(T), < \ln X_2(T) >)$	-0.4148	-0.4511	-0.4420	-0.4280

Table 3. First four moments for scenarios on 3/2 component (*b*), Scenario **B**.

	$b_1 = 0.008, \tilde{b}_1 = \tilde{b}_2 = 0$	$b_1 = 0, \tilde{b}_1 = \tilde{b}_2 = 0.008$	$b_1 = \tilde{b}_1 = \tilde{b}_2 = 0$	$b_1 = \tilde{b}_1 = \tilde{b}_2 = 0.008$
$\mathbb{E}\left[\frac{X_1(T)-X_1(0)}{X_1(0)}\right]$	0.0514	0.0504	0.0527	0.0499
$\mathbb{E}\left[\frac{X_2(T)-X_2(0)}{X_2(0)}\right]$	0.0774	0.0775	0.0787	0.0754
$\mathbb{V}\left[\frac{X_1(T)-X_1(0)}{X_1(0)}\right]$	0.0360	0.0606	0.0247	0.1022
$\mathbb{V}\left[\frac{X_2(T)-X_2(0)}{X_2(0)}\right]$	0.0359	0.1508	0.0248	0.0723
Corr ($\ln X_1(T), \ln X_2(T)$)	0.7533	0.0099	0.4698	0.0156
Corr ($\ln X_1(T), < \ln X_1(T) >$)	-0.4509	-0.2031	-0.5560	-0.0273
Corr ($\ln X_2(T), < \ln X_2(T) >$)	-0.4496	-0.0398	-0.5517	-0.2444

Similarly, we considered the following scenarios for a : $a_1 = a_2 = 0$; $a_1 = 0.75, a_2 = 0$; $a_1 = 0, a_2 = 0.75$ and $a_1 = a_2 = 0.75$. Tables 4 and 5 present key statistics for the returns under Scenarios **A** and **B**, respectively.

Table 4. First four moments for scenarios on commonalities (*a*), Scenario **A**.

	$a_1 = a_2 = 0$	$a_1 = 0.75, a_2 = 0$	$a_1 = 0, a_2 = 0.75$	$a_1 = a_2 = 0.75$
$\mathbb{E}\left[\frac{X_1(T)-X_1(0)}{X_1(0)}\right]$	0.0491	0.0492	0.0493	0.0492
$\mathbb{E}\left[\frac{X_2(T)-X_2(0)}{X_2(0)}\right]$	0.0761	0.0757	0.0746	0.0759
$\mathbb{V}\left[\frac{X_1(T)-X_1(0)}{X_1(0)}\right]$	0.0658	0.0727	0.0660	0.0732
$\mathbb{V}\left[\frac{X_2(T)-X_2(0)}{X_2(0)}\right]$	0.0371	0.0370	0.0444	0.0447
Corr ($\ln X_1(T), \ln X_2(T)$)	0.0000	-0.0011	0.0004	0.2841
Corr ($\ln X_1(T), < \ln X_1(T) >$)	-0.4523	-0.4400	-0.4529	-0.4430
Corr ($\ln X_2(T), < \ln X_2(T) >$)	-0.4507	-0.4521	-0.4285	-0.4279

Table 5. First four moments for scenarios on commonalities (*a*), Scenario **B**.

	$a_1 = a_2 = 0$	$a_1 = 0.75, a_2 = 0$	$a_1 = 0, a_2 = 0.75$	$a_1 = a_2 = 0.75$
$\mathbb{E}\left[\frac{X_1(T)-X_1(0)}{X_1(0)}\right]$	0.0514	0.0496	0.0512	0.0500
$\mathbb{E}\left[\frac{X_2(T)-X_2(0)}{X_2(0)}\right]$	0.0768	0.0772	0.0756	0.0752
$\mathbb{V}\left[\frac{X_1(T)-X_1(0)}{X_1(0)}\right]$	0.0420	0.0733	0.0420	0.0857
$\mathbb{V}\left[\frac{X_2(T)-X_2(0)}{X_2(0)}\right]$	0.0507	0.0564	0.0746	0.0719
Corr ($\ln X_1(T), \ln X_2(T)$)	0.0004	-0.0000	-0.0005	0.0039
Corr ($\ln X_1(T), < \ln X_1(T) >$)	-0.1126	-0.0047	-0.3203	-0.2634
Corr ($\ln X_2(T), < \ln X_2(T) >$)	-0.1600	-0.2544	-0.0164	-0.0111

4.1. Pricing Option

The section prices European call option on the asset X_1 based on our 4/2 generalized factor model. It explores the implied volatility surface in a three-dimensional plot with strike prices as the x-axis, time to maturity as the y-axis, and corresponding implied volatility as z-axis. We take the strike prices K to be 15, 16.4, 17.8, 19.2, 20.6, and 22 and the expiry dates T are 0.2, 0.36, 0.52, 0.68, 0.84 and 1.0. By choosing these strike prices, we account for the in-the-money, at-the-money, and out-of-the-money options, given the initial asset price 18. Subsequently, for each strike price and expiry date, we can obtain a simulated call price as follows

$$c(T, K) = e^{-rT} \mathbb{E}^Q[(X_1(T) - K)^+],$$

where $X_1(T)$ is approximated using the Euler method.

We extract the implied volatility by matching the Black–Scholes option price formula with simulated call prices and solve for the volatility parameter. Hence, we can treat the dynamics of $Y(t)$ as an O-U process such that:

$$dY(t) = (L_1 - 0.5\sigma^2 - \beta Y(t))dt + \sigma dW^*(t).$$

Next, we consider the two cases described above. The first one studies the impact of b , which represents the size of the 3/2 component on the covariance; and the second examines the impact of a , the size of the commonality.

In the case of b , we first extract the implied volatility surface by matching the standard BS formula for changes on b and \tilde{b} respectively (see Figures 1 and 2).

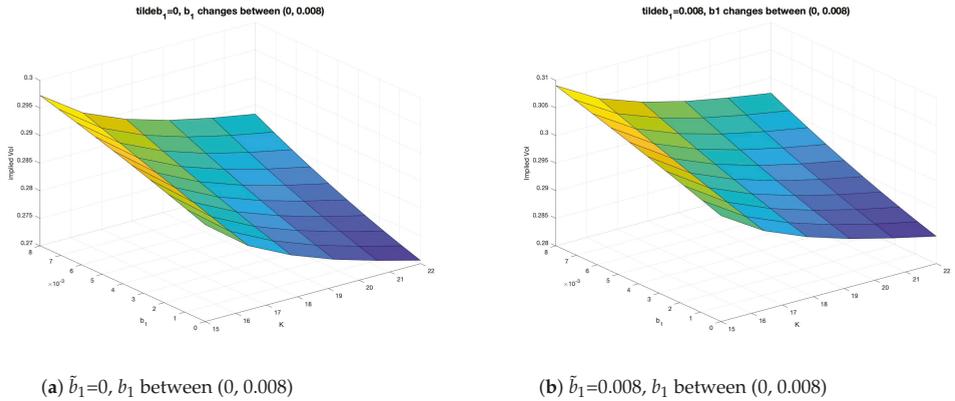


Figure 1. Impact of b_1 (common factor, 3/2 component) on implied volatility, Scenario A.

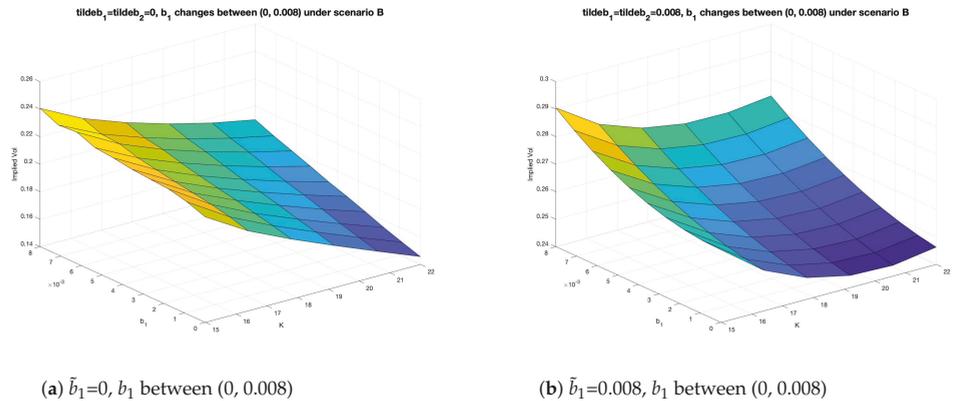


Figure 2. Impact of b_1 (common factor, 3/2 component) on implied volatility, Scenario B.

For Scenario A, Figure 1a,b illustrates that even small changes in (b_1) the common factor 3/2 component (from 0 to 0.008) can lead to a 7% difference in implied volatility (from 0.275 to 0.295, or 0.285 to 0.305). The joint effect of the common and intrinsic 3/2 components (b_1 and \tilde{b}_1) can be obtained by combining those two figures leading to a 11% change (from 0.275 to 0.305) in the presence of relatively small values of b .

For Scenario B, we observe that the impact of intrinsic factor on volatility surface is more significant than in Scenario A through a comparison of Figure 2a,b. The effect of b_1 on implied

volatility increase by approximately 31% (0.145 to 0.19), as shown in Figure 2a, when only the common factor is present. In Figure 2b, we observe a volatility “smile” with the difference of approximately 12.2% (0.245 to 0.275). The joint effect of the common and intrinsic 3/2 components in this case is 100% (0.145 to 0.29).

Figures 1 and 2 jointly demonstrate that, given different underlying process for common and intrinsic factors, the impact of the 3/2 component can be crucial.

Next, we study a , the weight of the common factor (commonality). We again extract the implied volatility surface from matching the standard BS formula for changes on a .

Figure 3a,b displays the significant increase in implied volatility due to the commonality of the asset with the market (a_1). The change in implied volatility can increase up to 12.5% (from 0.28 to 0.315) in Scenario A and up to 30% (from 0.22 to 0.32) in Scenario B.

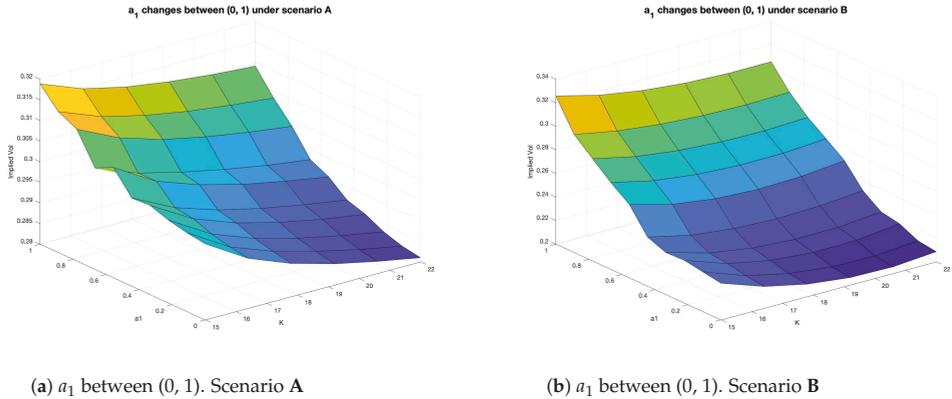


Figure 3. Impact of commonality (a_1) on implied volatility.

4.2. Risk Measures

This section examines the impact of b and a on important risk measures, in particular the value at risk (VaR) and the expected shortfall (ES). For clarity and calculation purposes, these measures are defined as follows:

$$\alpha = P(X(T) \leq -VaR_\alpha) \tag{10}$$

$$ES = -\frac{1}{\alpha} \int_0^\alpha VaR_\gamma d\gamma \tag{11}$$

where $X(T) = \omega_1(X_1(T) - X_1(0)) + \omega_2(X_2(T) - X_2(0))$ is the profit and loss portfolio with equal weights ($\omega_1 = \omega_2 = 1/2$) (see DeMiguel et al. 2007 for a rationale and support of this simple strategy). We let α vary from 0.001 to 0.2 with a discretization size of 200.

We first study the impact of b_1, \bar{b}_1 and \bar{b}_2 on VaR and ES for a fixed value of $\alpha = 0.01$. Figure 4a,b illustrates a substantial increase in VaR, from 16 (when all b values are set to zero) to 19.5 (all b set to 0.008), which is a 21% increase ($\alpha = 0.01$) due to the presence of b . In other words, an investor would have to set aside 21% more capital in the presence of 3/2 components. Similarly, ES increases from -21 in the presence of 3/2 components to -18.5 in their absence, which constitutes a 12% increase in the average VaR.

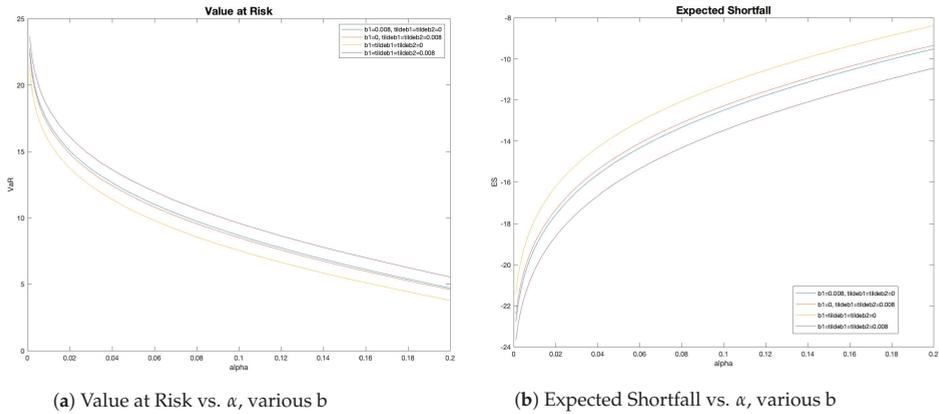


Figure 4. Impact of 3/2 components (*b*) on Risk measures, Scenario A.

Figure 5a,b also displays a substantial increase in VaR, from 17.5 (all *b* set to zero) to 22.5 (*b* set to 0.008), which represents a 28.6% increase ($\alpha = 0.01$) due to *b*. This means 28.6% more capital is required in the presence of 3/2 components. Similarly, the ES increases from -27 with 3/2 components to -19 without them, representing a 29.6% increase in the average VaR.

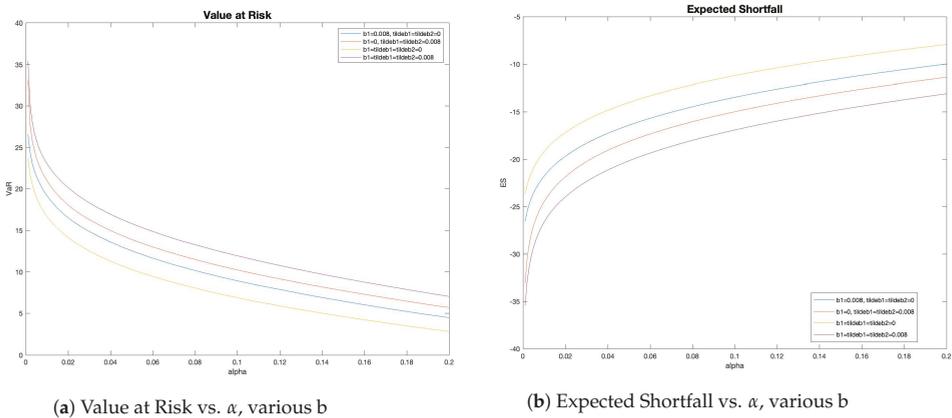
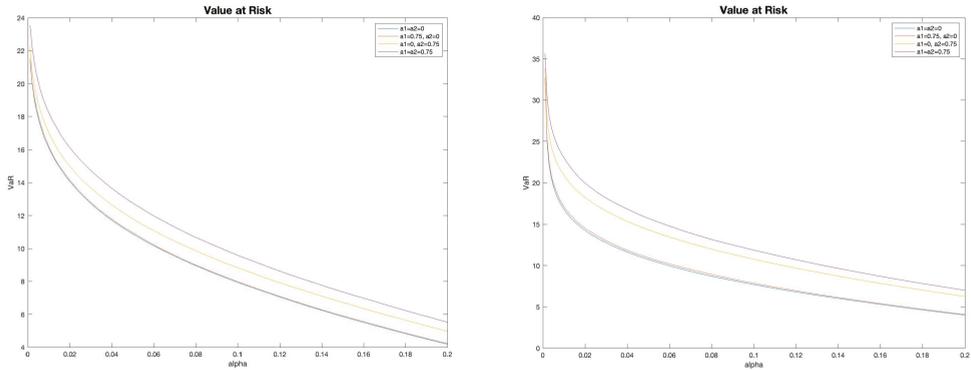


Figure 5. Impact of 3/2 components (*b*) on Risk measures, Scenario B.

A similar analysis was performed with respect to the commonality *a*, in the presence of stochastic volatility (in the common factor) versus constant volatility; in other words, we assessed the impact of *a* per se and that of stochastic correlation produced by the 4/2 model. Figure 6a demonstrates an increase in VaR, from 16 to 18.5, a 15.5% increase ($\alpha = 0.01$). Figure 6b shows that the VaR jumps from 17 to 23, a 35% growth due to the increase in *a*.



(a) Scenario A: Value at Risk vs. α for different commonality (a) values.

(b) Scenario B: Value at Risk vs. α for different commonality (a) values.

Figure 6. Impact of commonality (a) on Value at Risk under Scenario A and Scenario B.

5. Conclusions

The generalized mean-reverting 4/2 factor model is proposed and studied in this paper. We provide analytical expressions for key characteristic functions and conditions for well-defined changes of measures. We also explore the impact of b , i.e., the 3/2 component of the model in the volatility process, and a , the commonalities in the absence and presence of stochastic volatility on the common factor. These impacts were measured with respect to implied volatility surfaces and two important risk measures. The results demonstrate that even small values of the 3/2 component (b) can lead to a 100% change in the implied volatility surface, as well as up to 28% and 29% increases in the VaR and ES measures, respectively.

Author Contributions: All authors contributed equally.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proofs

Proof. Proof of Proposition 1.

The first step is to ensure the change of measure is well-defined and for this we use Novikov’s condition, i.e., generically

$$\mathbb{E} \left[\exp \left(\frac{1}{2} \int_0^T \lambda^2 \left(\sqrt{v(t)} + \frac{b}{\sqrt{v(t)}} \right)^2 ds \right) \right] = e^{\lambda^2 b T} \mathbb{E} \left[\exp \left(\frac{\lambda^2}{2} \int_0^T v(s) ds + \frac{\lambda^2 b^2}{2} \int_0^T \frac{1}{v(s)} ds \right) \right] < \infty.$$

From Grasselli, in order for this expectation to exist, we need two conditions:

$$-\frac{\lambda^2}{2} > -\frac{\alpha^2}{2\bar{\xi}^2} \implies |\lambda| < \frac{\alpha}{\bar{\xi}} \tag{A1}$$

and

$$-\frac{\lambda^2 b^2}{2} \geq -\frac{(2\alpha\theta - \bar{\xi}^2)^2}{8\bar{\xi}^2} \implies |\lambda| \leq \frac{2\alpha\theta - \bar{\xi}^2}{2|b|\bar{\xi}} \implies \bar{\xi}^2 \leq 2\alpha\theta - 2|\lambda||b|\bar{\xi} \tag{A2}$$

The latter condition in Equation (A2) implies, in particular, that our volatility processes satisfy Feller’s condition under \mathcal{P} and \mathcal{Q} ; in other words, it ensures all our CIR processes stay away from zero under both measures.

Applying Equation (A2) to our setting leads to $(i, j = 1, \dots, n)$:

$$\zeta_j^2 \leq 2\alpha_j\theta_j - 2\zeta_j \max \left\{ |\lambda_j b_j|, |\lambda_j^\perp b_j| \right\} \tag{A3}$$

$$\tilde{\zeta}_i^2 \leq 2\tilde{\alpha}_i\tilde{\theta}_i - 2\tilde{\zeta}_i \max \left\{ |\tilde{\lambda}_i \tilde{b}_i|, |\tilde{\lambda}_i^\perp \tilde{b}_i| \right\} \tag{A4}$$

Now, we apply Equation (A1) producing two extra set of conditions $(i, j = 1, \dots, n)$:

$$\max \left\{ |\lambda_j|, |\lambda_j^\perp| \right\} < \frac{\alpha_j}{\zeta_j} \tag{A5}$$

$$\max \left\{ |\tilde{\lambda}_i|, |\tilde{\lambda}_i^\perp| \right\} < \frac{\tilde{\alpha}_i}{\tilde{\zeta}_i} \tag{A6}$$

The second step applies to the case $\beta_{ij} = 0$ for $i, j = 1, \dots, n$ and it is to ensure the drift of the asset price equal the short rate:

$$L_i = r, c_i = \sum_{j=1}^n a_{ij} \left(\rho_j \lambda_j + \sqrt{1 - \rho_j^2} \lambda_j^\perp \right), \tilde{c}_i = \tilde{\rho}_i \tilde{\lambda}_i + \sqrt{1 - \tilde{\rho}_i^2} \tilde{\lambda}_i^\perp$$

For the most general case $(\beta_{ij} \neq 0$ for some i or j), the second step should be adapted to any particular prescribed drift structure under the Q -measure.

The third step is to ensure the drift-less asset price process is a true Q -martingale and not just a local Q -martingale:

$$\frac{dX_i(t)}{X_i(t)} = (\cdot) dt + \sum_{j=1}^n a_{ij} \left(\sqrt{v_j(t)} + \frac{b_j}{\sqrt{v_j(t)}} \right) dW_j^Q(t) + \left(\sqrt{\tilde{v}_i(t)} + \frac{\tilde{b}_i}{\sqrt{\tilde{v}_i(t)}} \right) d\tilde{W}_i^Q(t)$$

Here, we test the martingale property using the Feller nonexplosion test for volatilities, hence considering the following $n^2 + n$ changes of Brownian motion for the volatility processes and checking the processes do not reach zero under the various measures:

$$dB_{ij}^Q(t) = a_{ij}\rho_j \left(\sqrt{v_j(t)} + \frac{b_j}{\sqrt{v_j(t)}} \right) dt + dB_j^P(t), d\tilde{B}_i^Q(t) = \tilde{\rho}_i \left(\sqrt{\tilde{v}_i(t)} + \frac{\tilde{b}_i}{\sqrt{\tilde{v}_i(t)}} \right) dt + d\tilde{B}_i^P(t)$$

This leads to the following conditions:

$$\zeta_j^2 \leq 2\alpha_j\theta_j - 2 |a_{ij}\rho_j b_j| \zeta_j, i, j = 1, \dots, n \tag{A7}$$

$$\tilde{\zeta}_i^2 \leq 2\tilde{\alpha}_i\tilde{\theta}_i - 2 |\tilde{\lambda}_i \tilde{\rho}_i \tilde{b}_i| \tilde{\zeta}_i, i = 1, \dots, n \tag{A8}$$

We can combine the first and third steps in Equations (A3), (A7), (A4) and (2) into the final conditions. □

Proof. Proof of Proposition 2.

We start by defining new processes $dS_{ik,t}$ and dS_t^{ij} with $i, j, k = 1, 2, \dots, n$ and $t \geq 0$.

$$\begin{aligned} dZ_i(t) &= \sum_{j=1}^n \left(e^{\beta t} \right)_{ij} \left[\left(\tilde{c}_j - \frac{1}{2} \right) \left(\sqrt{\tilde{v}_j(t)} + \frac{\tilde{b}_j}{\sqrt{\tilde{v}_j(t)}} \right) dt + \left(\sqrt{\tilde{v}_j(t)} + \frac{\tilde{b}_j}{\sqrt{\tilde{v}_j(t)}} \right) d\tilde{W}_j \right] \\ &+ \sum_{k=1}^n \left[\left(\sum_{j=1}^n \left(e^{\beta t} \right)_{ij} \frac{L_j}{n} \right) dt + \left(\sum_{j=1}^n \left(e^{\beta t} \right)_{ij} \left(c_k - \frac{1}{2} \right) a_{jk}^2 \right) \left(\sqrt{v_k(t)} + \frac{b_k}{\sqrt{v_k(t)}} \right)^2 dt \right] \\ &+ \sum_{k=1}^n \left(\sum_{j=1}^n \left(e^{\beta t} \right)_{ij} a_{jk} \right) \left(\sqrt{v_k(t)} + \frac{b_k}{\sqrt{v_k(t)}} \right) dW_k \\ &= \sum_{j=1}^n dS_t^{ij} + \sum_{k=1}^n dS_{ik,t} \end{aligned}$$

By the dependence structure implied by the model, it follows that all S are independent for a fix i , hence we can transform the characteristic function using the processes S as follows:

$$\Phi_{Z(t),v(t)}(T, \omega) = \prod_{k=1}^n E \left[\exp \{ i \omega' (S_{\cdot,k,T} - S_{\cdot,k,t}) \} \mid S_t, v(t) \right] \prod_{j=1}^n E \left[\exp \{ i \omega_j (S_T^j - S_t^j) \} \mid S_t, v(t) \right]$$

For each factor $j = 1, 2, \dots, n$ we define $S_{k,t}^* = \omega' S_{\cdot,k,t} = \sum_{i=1}^n \omega_i S_{ik,t}$; the dynamics of $S_{k,t}^*$ can be expressed as

$$\begin{aligned} dS_{k,t}^* &= \omega' dS_{\cdot,k,t} \\ &= \left[L(\omega, t) + h_k(\omega, t) \left(\sqrt{v_k(t)} + \frac{b_k}{\sqrt{v_k(t)}} \right)^2 \right] dt + g_k(\omega, t) \left(\sqrt{v_k(t)} + \frac{b_k}{\sqrt{v_k(t)}} \right) dW_{k,t} \end{aligned}$$

where $h_k(\omega, t) = \sum_{j=1}^n \left(c_k - \frac{1}{2} \right) a_{jk}^2 f_j(\omega, t)$, $L(\omega, t) = \sum_{j=1}^n \frac{L_j}{n} f_j(\omega, t)$ and $g_k(\omega, t) = \sum_{j=1}^n a_{jk} f_j(\omega, t)$ and $f_j(\omega, t) = \sum_{i=1}^n \omega_i \left(e^{\beta t} \right)_{ij}$. These three functions are deterministic, linear combinations of $f_j(\omega, t)$.

Next, we find the characteristic function for the increments of $S_{k,t}^*$:

$$E \left[\exp \{ i \phi \left(S_{k,T}^* - S_{k,t}^* \right) \} \mid S_{k,t}^*, v_k(t) = v_{k,t} \right] = \Phi_{GG} \left(T, \phi; L(\omega), h_k(\omega), g_k(\omega), \kappa_k, \theta_k, \zeta_k, \rho_k, b_k, c_k, v_{k,t}, S_{k,t}^* \right).$$

The generic function Φ_{GG} is provided in Lemma A1.

It follows similarly for idiosyncratic factors:

$$\begin{aligned} dS_t^{*j} &= \omega' dS_t^j \\ &= \left(\sum_{i=1}^n \omega_i \left(\sum_{j=1}^n \left(e^{\beta t} \right)_{ij} \left(\tilde{c}_j - \frac{1}{2} \right) \right) \right) \left(\sqrt{\tilde{v}_j(t)} + \frac{\tilde{b}_j}{\sqrt{\tilde{v}_j(t)}} \right) dt \\ &+ \left(\sum_{i=1}^n \omega_i \sum_{j=1}^n \left(e^{\beta t} \right)_{ij} \right) \left(\sqrt{\tilde{v}_j(t)} + \frac{\tilde{b}_j}{\sqrt{\tilde{v}_j(t)}} \right) d\tilde{W}_{j,t} \\ &= h_j(\omega, t) \left(\sqrt{\tilde{v}_j(t)} + \frac{\tilde{b}_j}{\sqrt{\tilde{v}_j(t)}} \right)^2 dt + g_j(\omega, t) \left(\sqrt{\tilde{v}_j(t)} + \frac{\tilde{b}_j}{\sqrt{\tilde{v}_j(t)}} \right) d\tilde{W}_{j,t} \end{aligned}$$

where $h_j(\omega, t) = \sum_{i=1}^n \omega_i \left(\sum_{j=1}^n \left(e^{\beta t} \right)_{ij} \left(\tilde{c}_j - \frac{1}{2} \right) \right)$ and $g_j(\omega, t) = \sum_{i=1}^n \omega_i \sum_{j=1}^n \left(e^{\beta t} \right)_{ij}$.

Combining all pieces together, we obtain:

$$\begin{aligned} \Phi_{Z(t),v(t)}(T, \omega) &= \prod_{k=1}^n E \left[\exp\{i\omega' (S_{k,T} - S_{k,t})\} \mid S_t, v(t) \right] \prod_{j=1}^n E \left[\exp\{i\omega' (S_T^j - S_t^j)\} \mid S_t, v(t) \right] \\ &= \prod_{k=1}^n \Phi_{GG} \left(T, 1; L_k(\omega), h_k(\omega), g_k(\omega), \kappa_k, \theta_k, \xi_k, \rho_k, b_k, c_k, v_{k,t}, S_{k,t}^* \right) \\ &\times \prod_{j=1}^n \Phi_{GG} \left(T, 1; 0, h_j(\omega), g_j(\omega), \tilde{\kappa}_j, \tilde{\theta}_j, \tilde{\xi}_j, \tilde{\rho}_j, \tilde{b}_j, \tilde{c}_j, \tilde{v}_{j,t}, S_t^{*j} \right) \end{aligned}$$

□

Lemma A1. Let the generic process be:

$$\begin{aligned} dZ(t) &= \left[L(t) + h(t) \left(\sqrt{v(t)} + \frac{b}{\sqrt{v(t)}} \right)^2 \right] dt + g(t) \left(\sqrt{v(t)} + \frac{b}{\sqrt{v(t)}} \right) dW_t \\ dv(t) &= \alpha(\theta - v(t))dt + \xi \sqrt{v(t)} dB(t) \\ \langle dB(t), dW(t) \rangle &= \rho dt \end{aligned}$$

with $g(t)$ differentiable, then

$$\begin{aligned} \Phi_{GG}(T, \phi; L, h, g, \kappa, \theta, \xi, \rho, b, c, v_t, Z_t) &= \exp \left\{ u \int_t^T A(s) ds \right\} v(t)^{u \frac{b\rho}{\xi}} \exp \left\{ -u\rho \frac{g(t)v(t)}{\xi} \right\} \\ \times E \left[v(T)^{u \frac{b\rho}{\xi}} \exp \left\{ u \left(\int_t^T B(s)v(s)ds + \int_t^T C(s) \frac{1}{v(s)} ds + \int_t^T D(s) \ln(v(s))ds + \rho \frac{g(T)v(T)}{\xi} \right) \right\} \mid \mathcal{F}_t \right] \end{aligned}$$

where A, B, C and D are provided in the proof.

Proof. Proof of Lemma A1.

Let the generic process be:

$$\begin{aligned} dZ(t) &= \left[L(t) + h(t) \left(\sqrt{v(t)} + \frac{b}{\sqrt{v(t)}} \right)^2 \right] dt + g(t) \left(\sqrt{v(t)} + \frac{b}{\sqrt{v(t)}} \right) dW_t \\ dv(t) &= \alpha(\theta - v(t))dt + \xi \sqrt{v(t)} dB(t) \\ \langle dB(t), dW(t) \rangle &= \rho dt \end{aligned}$$

We want to find

$$\mathbb{E} \left[e^{uZ(T)} \mid \mathcal{F}_t \right] = e^{uZ(t)} \Phi_{GG}(T, \phi; L, h, g, \kappa, \theta, \xi, \rho, b, c, v_t, Z_t)$$

Letting $\tilde{v}(t) = g(t)v(t)$ and $\hat{v}(t) = g(t)\ln(v(t))$, we have following:

$$d\tilde{v}(t) = \alpha\theta g(t) dt + (g'(t) - \alpha g(t))v(t)dt + g(t)\xi\sqrt{v(t)}dB(t), \tag{A9}$$

and

$$\begin{aligned} d\hat{v}(t) &= \frac{g'(t)}{g(t)}\hat{v}(t)dt + \frac{\partial\hat{v}(t)}{\partial\ln(v(t))}d\ln(v(t)) + \frac{1}{2}\frac{\partial^2\hat{v}(t)}{\partial\ln(v(t))^2} < d\ln(v(t)) > \\ &= \frac{g'(t)}{g(t)}\hat{v}(t)dt + \frac{g(t)\xi}{\sqrt{v(t)}}dB(t) + g(t)\left(\frac{\alpha\theta}{v(t)} - \alpha\right)dt - g(t)\frac{\xi^2}{2v(t)}dt \end{aligned} \tag{A10}$$

From Equations (A9) and (A10), we solve for $\int_t^T g(s) \sqrt{v(s)} dB(s)$ and $\int_t^T \frac{g(s)}{\sqrt{v(s)}} dB(s)$:

$$\int_t^T g(s) \sqrt{v(s)} dB(s) = \frac{g(T)v(T) - g(t)v(t)}{\xi} - \frac{\alpha\theta}{\xi} \int_t^T g(s) ds - \frac{1}{\xi} \int_t^T (g'(s) - \alpha g(s)) v(s) ds, \tag{A11}$$

$$\int_t^T \frac{g(s)}{\sqrt{v(s)}} dB(s) = \frac{1}{\xi} \ln \frac{v(T)g(T)}{v(t)g(t)} + \frac{\alpha}{\xi} \int_t^T g(s) ds + \frac{1}{\xi} \left(\frac{\xi^2}{2} - \alpha\theta \right) \int_t^T \frac{g(s)}{v(s)} ds - \frac{1}{\xi} \int_t^T g'(s) \ln(v(s)) ds. \tag{A12}$$

Split $W(t)$ into $B(t)$ and its orthogonal part $B(t)^\perp$:

$$\begin{aligned} Z(T) = Z(t) &+ \int_t^T L(s) ds + \int_t^T h(s) \left(\sqrt{v(s)} + \frac{b}{\sqrt{v(s)}} \right)^2 ds \\ &+ \int_t^T g(s) \left(\sqrt{v(s)} + \frac{b}{\sqrt{v(s)}} \right) (\rho dB(s) + \sqrt{1 - \rho^2} dB(s)^\perp), \end{aligned}$$

then substitute Equation (A11) and (A12) to eliminate $dB(t)$. $Z(t)$ can be rewritten now as:

$$\begin{aligned} Z(T) = Z(t) &+ \int_t^T L(s) ds + \int_t^T h(s) \left(v(s) + 2b + \frac{b^2}{v(s)} \right) ds + \sqrt{1 - \rho^2} \int_t^T g(s) \left(a\sqrt{v(s)} + \frac{b}{\sqrt{v(s)}} \right) dB(s)^\perp \\ &+ \rho \frac{g(T)v(T) - g(t)v(t)}{\xi} - \frac{\alpha\theta\rho}{\xi} \int_t^T g(s) ds - \frac{\rho}{\xi} \int_t^T (g'(s) - \alpha g(s)) v(s) ds \\ &+ \frac{b\rho}{\xi} \ln \frac{v(T)g(T)}{v(t)g(t)} + \frac{\alpha b\rho}{\xi} \int_t^T g(s) ds + \frac{b\rho}{\xi} \left(\frac{\xi^2}{2} - \alpha\theta \right) \int_t^T \frac{g(s)}{v(s)} ds - \frac{b\rho}{\xi} \int_t^T g'(s) \ln(v(s)) ds \end{aligned}$$

Grouping conveniently, we obtain:

$$\begin{aligned} Z(T) = Z(t) &+ \int_t^T \bar{A}(s) ds + \int_t^T \bar{B}(s)v(s) ds + \int_t^T \bar{C}(s) \frac{1}{v(s)} ds + \int_t^T D(s) \ln(v(s)) ds \\ &+ \frac{b\rho}{\xi} \ln \frac{v(T)g(T)}{v(t)g(t)} + \rho \frac{g(T)v(T) - g(t)v(t)}{\xi} + \sqrt{1 - \rho^2} \int_t^T g(s) \left(a\sqrt{v(s)} + \frac{b}{\sqrt{v(s)}} \right) dB(s)^\perp \end{aligned}$$

where

$$\begin{aligned} \bar{A}(s) &= L(s) + 2bh(s) + \left(\frac{\alpha b\rho}{\xi} - \frac{\alpha\theta\rho}{\xi} \right) g(s) \\ \bar{B}(s) &= h(s) - \frac{\rho}{\xi} (g'(s) - \alpha g(s)) \\ \bar{C}(s) &= b^2h(s) + \frac{b\rho}{\xi} \left(\frac{\xi^2}{2} - \alpha\theta \right) g(s) \\ D(s) &= -\frac{b\rho}{\xi} g'(s) \end{aligned}$$

Let $(\mathcal{G}_t)_{t \geq 0}$ denote the filtration generated by $v(t), t \geq 0$. Using iterated expectation and independence, we can write the conditional moment generating function of $Z(T)$ as:

$$\begin{aligned} \mathbb{E} \left[e^{uZ(T)} | \mathcal{F}_t \right] &= \mathbb{E} \left[\mathbb{E} \left[e^{uZ(T)} | \mathcal{F}_t \cup \mathcal{G}_t \right] | \mathcal{F}_t \right] = \exp \left\{ u \left(Z(t) + \int_t^T A(s) ds \right) \right\} v(t)^{u \frac{b\rho}{\xi} g(t)} \exp \left\{ -u\rho \frac{g(t)v(t)}{\xi} \right\} \\ &\times \mathbb{E} \left[\exp \left\{ u \left(\int_t^T B(s)v(s) ds + \int_t^T C(s) \frac{1}{v(s)} ds + \int_t^T D(s) \ln(v(s)) ds + \frac{b\rho}{\xi} \ln v(T)g(T) + \rho \frac{g(T)v(T)}{\xi} \right) \right\} \right. \\ &\quad \left. \times \mathbb{E} \left[\exp \left\{ u \sqrt{1 - \rho^2} \int_t^T g(s) \left(\sqrt{v(s)} + \frac{b}{\sqrt{v(s)}} \right) dB(s)^\perp \right\} | \mathcal{F}_t \cup \mathcal{G}_t \right] | \mathcal{F}_t \right] \end{aligned}$$

The inner expectation, conditioned on $\mathcal{F}_t \cup \mathcal{G}_t$, leads to a normal random variable with mean 0 and variance (Ito's Isometry) $u^2(1 - \rho^2) \int_t^T g^2(s)(v(s) + \frac{b^2}{v(s)} + 2b)ds$. Putting all together:

$$\begin{aligned} \mathbb{E} \left[e^{uZ(T)} | \mathcal{F}_t \right] &= \exp \left\{ u \left(Z(t) + \int_t^T A(s) ds \right) \right\} v(t)^{u \frac{b\rho}{\xi} g(t)} \exp \left\{ -u\rho \frac{g(t)v(t)}{\xi} \right\} \\ &\times \mathbb{E} \left[v(T)^{u \frac{b\rho}{\xi} g(T)} \exp \left\{ u \left(\int_t^T B(s)v(s)ds + \int_t^T C(s) \frac{1}{v(s)} ds + \int_t^T D(s) \ln(v(s))ds + \rho \frac{g(T)v(T)}{\xi} \right) \right\} | \mathcal{F}_t \right] \end{aligned}$$

where

$$\begin{aligned} A(s) &= \bar{A}(s) + u(1 - \rho^2)bg^2(s) \\ B(s) &= \bar{B}(s) + \frac{1}{2}u^2(1 - \rho^2)g^2(s) \\ C(s) &= \bar{C}(s) + \frac{1}{2}u^2(1 - \rho^2)b^2g^2(s) \end{aligned}$$

□

Proof. Proof of Corollary 1.

The proof starts similarly to Proposition 2. We start by defining new processes $dS_{ij,t}$ and dS_t^i with $i, j = 1, 2, \dots, n$ and $t \geq 0$.

$$\begin{aligned} dZ_i(t) &= \left[\left(\tilde{c}_i - \frac{1}{2} \right) \left(\sqrt{\tilde{v}_i(t)} + \frac{\tilde{b}_i}{\sqrt{\tilde{v}_i(t)}} \right) dt + \left(\sqrt{\tilde{v}_i(t)} + \frac{\tilde{b}_i}{\sqrt{\tilde{v}_i(t)}} \right) d\tilde{W}_i \right] \\ &+ \sum_{j=1}^n \left[\frac{L_j}{n} dt + \left(c_j - \frac{1}{2} \right) a_{ij}^2 \left(\sqrt{v_j(t)} + \frac{b_j}{\sqrt{v_j(t)}} \right)^2 dt + a_{ij} \left(\sqrt{v_j(t)} + \frac{b_j}{\sqrt{v_j(t)}} \right) dW_j \right] \\ &= dS_t^i + \sum_{j=1}^n dS_{ij,t} \end{aligned}$$

By the dependence structure implied by the model, it follows that all S are independent for a fix i , hence we can transform the characteristic function using the processes S as follows:

$$\Phi_{Z(t),v(t)}(T, \omega) = \prod_{j=1}^n E \left[\exp \{ i\omega' (S_{j,T} - S_{j,t}) \} | S_t, v(t) \right] \prod_{i=1}^n E \left[\exp \{ i\omega_i (S_T^i - S_t^i) \} | S_t, v(t) \right]$$

For each factor $j = 1, 2, \dots, n$ we define $S_{j,t}^* = \omega' S_{j,t} = \sum_{i=1}^n \omega_i S_{ij,t}$, the dynamics of $S_{j,t}^*$ can be expressed as

$$\begin{aligned} dS_{j,t}^* &= \omega' dS_{j,t} \\ &= \left[L(\omega) + h_j(\omega) \left(\sqrt{v_j(t)} + \frac{b_j}{\sqrt{v_j(t)}} \right)^2 \right] dt + g_j(\omega) \left(\sqrt{v_j(t)} + \frac{b_k}{\sqrt{v_j(t)}} \right) dW_{j,t} \end{aligned}$$

where $h_j(\omega) = \sum_{i=1}^n \omega_i \left(c_i - \frac{1}{2} \right) a_{ij}^2$, $L(\omega) = \sum_{i=1}^n \omega_i \frac{L_i}{n}$ and $g_j(\omega) = \sum_{i=1}^n \omega_i a_{ij}$. Next, we find the characteristic function for the increments of $S_{j,t}^*$:

$$E \left[\exp \{ i\phi (S_{j,T}^* - S_{j,t}^*) \} | S_{j,t}^*, v_j(t) = v_{j,t} \right] = \Phi_G \left(T, \phi; L, h_j, g_j, \kappa_j, \theta_j, \xi_j, \rho_j, b_j, c_j, v_{j,t}, S_{j,t}^* \right).$$

The generic function Φ_G is provided in the Appendix B.

Similarly for the idiosyncratic factors dS_i^i with $i = 1, 2, \dots, n$:

$$dS_i^{*i} = \omega_i dS_i^i = h_i(\omega) \left(\sqrt{\tilde{v}_i(t)} + \frac{\tilde{b}_i}{\sqrt{\tilde{v}_i(t)}} \right)^2 dt + g_i(\omega) \left(\sqrt{\tilde{v}_i(t)} + \frac{\tilde{b}_i}{\sqrt{\tilde{v}_i(t)}} \right) d\tilde{W}_{i,t}$$

where $h_i(\omega) = \omega_i \left(\tilde{c}_i - \frac{1}{2} \right)$ and $g_i(\omega) = \omega_i$.

Combining all pieces together, we obtain:

$$\begin{aligned} \Phi_{Z(t),v(t)}(T, \omega) &= \prod_{j=1}^n E \left[\exp\{i\omega' (S_{j,T} - S_{j,t})\} \mid S_t, v(t) \right] \prod_{i=1}^n E \left[\exp\{i\omega_i (S_T^i - S_t^i)\} \mid S_t, v(t) \right] \\ &= \prod_{j=1}^n \Phi_G \left(T, \phi; L, h_j, g_j, \kappa_j, \theta_j, \xi_j, \rho_j, b_j, c_j, v_{j,t}, S_{j,t}^* \right) \times \prod_{i=1}^n \Phi_G \left(T, 1; 0, h_i, g_i, \tilde{\kappa}_i, \tilde{\theta}_i, \tilde{\xi}_i, \tilde{\rho}_i, \tilde{b}_i, \tilde{c}_i, \tilde{v}_{i,t}, S_t^{*i} \right) \end{aligned}$$

□

Proof. Proof of Corollary 2.

The proof uses Corollary 1, where we express the joint c.f. as the product of one dimensional c.f.s of 4/2 type.

$$\begin{aligned} \Phi_{Z(t),v(t)}(T, \omega) &= \prod_{j=1}^n E \left[\exp\{i\omega' (S_{j,T} - S_{j,t})\} \mid S_t, v(t) \right] \prod_{i=1}^n E \left[\exp\{i\omega_i (S_T^i - S_t^i)\} \mid S_t, v(t) \right] \\ &= \prod_{j=1}^n \Phi_G \left(T, \phi; L, h_j, g_j, \kappa_j, \theta_j, \xi_j, \rho_j, b_j, c_j, v_{j,t}, S_{j,t}^* \right) \times \prod_{i=1}^n \Phi_G \left(T, 1; 0, h_i, g_i, \tilde{\kappa}_i, \tilde{\theta}_i, \tilde{\xi}_i, \tilde{\rho}_i, \tilde{b}_i, \tilde{c}_i, \tilde{v}_{i,t}, S_t^{*i} \right) \end{aligned}$$

Hence, every one of these functions $(\Phi_G(T, \phi; L, h, g, \kappa, \theta, \xi, \rho, b, c, v_t, Z_t) = \mathbb{E} [e^{uZ(T)} \mid \mathcal{F}_t])$ capture the c.f. of a process of the type:

$$\begin{aligned} dZ(t) &= \left[L + h \left(\sqrt{v(t)} + \frac{b}{\sqrt{v(t)}} \right)^2 \right] dt + g \left(\sqrt{v(t)} + \frac{b}{\sqrt{v(t)}} \right) dW_t \\ dv(t) &= \alpha(\theta - v(t))dt + \xi \sqrt{v(t)} dB(t) \\ \langle dB(t), dW(t) \rangle &= \rho dt \end{aligned}$$

It is not difficult to realize therefore that the c.f. given $v(T)$ can be similarly computed for every one of those processes, hence one can infer:

$$\begin{aligned} \Phi_{Z(t),v(T)}(T, \omega) &= \prod_{j=1}^n \Phi_{G,T} \left(T, \phi; L, h_j, g_j, \kappa_j, \theta_j, \xi_j, \rho_j, b_j, c_j, v_{j,T}, S_{j,T}^* \right) \\ &\quad \times \prod_{i=1}^n \Phi_{G,T} \left(T, 1; 0, h_i, g_i, \tilde{\kappa}_i, \tilde{\theta}_i, \tilde{\xi}_i, \tilde{\rho}_i, \tilde{b}_i, \tilde{c}_i, \tilde{v}_{i,T}, S_{i,T}^{*i} \right) \end{aligned}$$

where $\Phi_{G,T}(T, \phi; L, h, g, \kappa, \theta, \xi, \rho, b, c, v_T, Z_t) = \mathbb{E} [e^{uZ(T)} \mid \mathcal{F}_t \cup v(T)]$ is provided next in Appendix B. □

Appendix B. Helpful Results

Given the 4/2 process, the following c.f. are used in this paper:

$$dZ(t) = \left[L + h \left(\sqrt{v(t)} + \frac{b}{\sqrt{v(t)}} \right)^2 \right] dt + g \left(\sqrt{v(t)} + \frac{b}{\sqrt{v(t)}} \right) dW_t$$

$$dv(t) = \alpha(\theta - v(t))dt + \xi \sqrt{v(t)}dB(t)$$

$$\langle dB(t), dW(t) \rangle = \rho dt$$

•

$$\begin{aligned} \Phi_G(T, u; L, h, g, \alpha, \theta, \xi, \rho, b, c, v_t, Z_t) &= \mathbb{E} \left[e^{uZ(T)} | \mathcal{F}_t \right] \\ &= \exp \left\{ uZ(t) + \frac{\alpha^2 \theta}{\xi^2} (T-t) + u \left(r + 2(h - \frac{1}{2})g^2b - \frac{g\rho\alpha\theta}{\xi} + \frac{gb\rho\alpha}{\xi} \right) (T-t) + u^2(1-\rho^2)g^2b(T-t) \right\} \\ &\times \left(\frac{\sqrt{A_u}}{\xi^2 \sinh(\frac{\sqrt{A_u}}{2}t)} \right)^{m_u+1} v(t)^{\frac{1}{2} + \frac{m_u}{2} - \frac{\alpha\theta}{\xi^2} - \frac{ugbp}{\xi}} \left(K_u(T) - \frac{ug\rho}{\xi} \right)^{-\left(\frac{1}{2} + \frac{m_u}{2} + \frac{\alpha\theta}{\xi^2} + \frac{ugbp}{\xi}\right)} \\ &\times \exp \left\{ \frac{v(t)}{\xi^2} \left(-\sqrt{A_u} \coth \left(\frac{\sqrt{A_u}(T-t)}{2} \right) + \alpha - ug\rho\xi \right) \right\} \frac{\Gamma \left(\frac{1}{2} + \frac{m_u}{2} + \frac{\alpha\theta}{\xi^2} + \frac{ugbp}{\xi} \right)}{\Gamma(m_u + 1)} \\ &\times {}_1F_1 \left(\frac{1}{2} + \frac{m_u}{2} + \frac{\alpha\theta}{\xi^2} + \frac{ugbp}{\xi}, m_u + 1, \frac{A_u v(t)}{\xi^4 \left(K_u(T) - \frac{ug\rho}{\xi} \right) \sinh^2 \left(\frac{\sqrt{A_u}(T-t)}{2} \right)} \right), \end{aligned}$$

with

$$A_u = \alpha^2 - 2\xi^2 \left(u \left(\frac{g\rho\alpha}{\xi} + (h - \frac{1}{2})g^2 \right) + \frac{1}{2}u^2(1-\rho^2)g^2 \right),$$

$$m_u = \frac{2}{\xi^2} \sqrt{ \left(\alpha\theta - \frac{\xi^2}{2} \right)^2 - 2\xi^2 \left(u \left(\frac{gb\rho}{\xi} \left(\frac{\xi^2}{2} - \alpha\theta \right) + (h - \frac{1}{2})g^2b^2 \right) + \frac{1}{2}u^2(1-\rho^2)g^2b^2 \right) },$$

$$K_u(T) = \frac{1}{\xi^2} \left(\sqrt{A_u} \coth \left(\frac{\sqrt{A_u}(T-t)}{2} \right) + \alpha \right)$$

•

$$\begin{aligned} \Phi_{G,T}(T, u; L, h, g, \kappa, \theta, \xi, \rho, b, c, v_t, Z_t) &= \mathbb{E} \left[e^{uZ(T)} | \mathcal{F}_t \cup v(T) \right] \\ &= \exp \left\{ uZ(t) + u \left(r + 2(h - \frac{1}{2})g^2b - \frac{a\rho\alpha\theta}{\xi} + \frac{b\rho\alpha}{\xi} \right) (T-t) + u^2(1-\rho^2)g^2b(T-t) \right\} \\ &\times \exp \left\{ \frac{ug\rho}{\xi} (v(T) - v(t)) + \frac{ugbp}{\xi} \log \frac{v(T)}{v(t)} \right\} \\ &\times \frac{\sqrt{A_u} \sinh \left(\frac{\alpha(T-t)}{2} \right)}{\alpha \sinh \left(\frac{\sqrt{A_u}(T-t)}{2} \right)} \exp \left(\frac{v(T) + v(t)}{\xi^2} \left(\alpha \coth \left(\frac{\alpha(T-t)}{2} \right) - \sqrt{A_u} \coth \left(\frac{\sqrt{A_u}(T-t)}{2} \right) \right) \right) \\ &\times \frac{I_{\frac{2\alpha\theta}{\xi^2}} \sqrt{ \left(\alpha\theta - \frac{\xi^2}{2} \right)^2 + 2\xi^2 B_u } \left(\frac{2\sqrt{A_u}v(T)v(t)}{\xi^2 \sinh \left(\frac{\sqrt{A_u}(T-t)}{2} \right)} \right)}{I_{\frac{2\alpha\theta}{\xi^2}} - 1 \left(\frac{2\alpha \sqrt{v(T)v(t)}}{\xi^2 \sinh \left(\frac{\alpha(T-t)}{2} \right)} \right)}, \end{aligned}$$

with

$$B_u = u \left(\frac{gb\rho}{\xi} \left(\frac{\xi^2}{2} - \alpha\theta \right) + (h - \frac{1}{2})g^2b^2 \right) + \frac{1}{2}u^2(1-\rho^2)g^2b^2,$$

References

- Baldeaux Jan, Martino Grasselli, and Eckhard Platen. 2015. Pricing Currency Derivatives Under the Benchmark Approach. *Journal of Banking & Finance* 53: 34–48.
- Carr, Peter, and Dilip B. Madan. 1999. Option Valuation using the FFT. *Journal of Computational Finance* 2: 61–73. [[CrossRef](#)]
- Caldana, Ruggero, and Gianluca Fusai. 2013. A general closed-form spread option pricing formula. *Journal of Banking and Finance* 37: 4893–906. [[CrossRef](#)]
- Cheng, Peng, and Olivier Scaillet. 2007. Linear-quadratic jump-diffusion modelling. *Mathematical Finance* 17: 575–98. [[CrossRef](#)]
- Chiarella, Carl, Boda Kang, Christina Sklibosios Nikitopoulos, and Thuy Tô. 2013. Humps in the volatility structure of the crude oil futures market: New evidence. *Energy Economics* 40: 989–1000. [[CrossRef](#)]
- Christoffersen, Peter, Steven Heston, and Kris Jacobs. 2009. The shape and term structure of the index option smirk: Why multifactor stochastic volatility models work so well. *Management Science*. 55: 1914–32. [[CrossRef](#)]
- De Col, Alvise, Alessandro Gnoatto, and Martino Grasselli. 2013. Smiles all around: FX joint calibration in a multi-Heston model. *Journal of Banking & Finance* 37: 3799–818.
- Da Fonseca, José, Martino Grasselli, and Claudio Tebaldi. 2007. Option Pricing when Correlations are Stochastic: An Analytical Framework. *Review of Derivatives Research* 10: 151–80. [[CrossRef](#)]
- DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal. 2007. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy? *The Review of Financial Studies* 22: 1915–53. [[CrossRef](#)]
- Diebold, Francis X., and Marc Nerlove. 1989. The dynamics of exchange rate volatility: A multivariate latent factor ARCH model. *Journal of Applied Econometrics* 4: 1–21. [[CrossRef](#)]
- Escobar, Marcos. 2018. A Stochastic Volatility Factor Model of Heston type. Statistical Properties and Estimation. *Stochastics* 90: 172–99. [[CrossRef](#)]
- Escobar, Marcos, and Christoph Gschnaidtner. 2016. Parameters recovery via calibration in the Heston model: A comprehensive review. *Wilmott* 2016: 60–81. [[CrossRef](#)]
- Escobar, Marcos, and Zhenxian Gong. 2019. The mean-reverting 4/2 stochastic volatility model: Properties and Financial Applications. submitted.
- Engle, Robert. 2002. Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models. *Journal of Business & Economic Statistics* 20: 339–50.
- Fusai, Gianluca, Marina Marena, and Andrea Roncoroni. 2018. Analytical pricing of discretely monitored Asian-style options: Theory and application to commodity markets. *Journal of Banking and Finance* 32: 2033–45. [[CrossRef](#)]
- Gourieroux, Christian. 2006. Continuous Time Wishart Process for Stochastic Risk. *Econometric Reviews* 25: 1–41. [[CrossRef](#)]
- Grasselli, Martino. 2017. The 4/2 stochastic volatility model: A unified approach for the Heston and the 3/2 model. *Mathematical Finance* 27: 1013–34. [[CrossRef](#)]
- Gnoatto, Alessandro, Martino Grasselli, and Eckhard Platen. 2018. A penny saved is a penny earned: Less expensive zero coupon bonds. *arXiv*, arXiv:1608.04683.
- Heston, Steven L. 1993. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bonds and Currency Options. *Review of Financial Studies* 6: 327–43. [[CrossRef](#)]
- Medvedev Alexey, and Olivier Scaillet. 2007. Approximation and calibration of short-term implied volatilities under jump-diffusion stochastic volatility. *Review of Financial Studies* 20: 427–59. [[CrossRef](#)]
- Muhle-Karbe, Johannes, Oliver Pfaffel, and Robert Stelzer. 2012. Option pricing in multivariate stochastic volatility models of OU type. *SIAM Journal on Financial Mathematics* 3: 66–94. [[CrossRef](#)]
- Platen, Eckhard. 1997. *A Non-Linear Stochastic Volatility Model*. Financial Mathematics Research Report No. FMRR 005-97. Canberra: Center for Financial Mathematics, Australian National University.
- Platen, Eckhard, and David Cochran Heath. 2010. *A Benchmark Approach to Quantitative Finance*. Berlin and Heidelberg: Springer.

- Schneider Lorenz, and Bertrand Tavin. 2018. From the Samuelson volatility effect to a Samuelson correlation effect: An analysis of crude oil calendar spread options. *Journal of Banking and Finance* 95: 185–202. [[CrossRef](#)]
- Schwartz, Eduardo S. 1997. The stochastic behavior of commodity prices: Implications for valuation and hedging. *The Journal of Finance* 52: 923–73. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Article

Defined Contribution Pension Plans: Who Has Seen the Risk?

Peter A. Forsyth ^{1,*} and Kenneth R. Vetzal ²

¹ David R. Cheriton School of Computer Science, University of Waterloo, Waterloo, ON N2L 3G1, Canada

² School of Accounting and Finance, University of Waterloo, Waterloo, ON N2L 3G1, Canada; kvetzal@uwaterloo.ca

* Correspondence: paforsyt@uwaterloo.ca; Tel.: +1-519-888-4567 (ext. 34415)

† This work was supported by the Natural Sciences and Engineering Research Council of Canada.

Received: 20 February 2019; Accepted: 18 April 2019; Published: 24 April 2019

Abstract: The trend towards eliminating defined benefit (DB) pension plans in favour of defined contribution (DC) plans implies that increasing numbers of pension plan participants will bear the risk that final realized portfolio values may be insufficient to fund desired retirement cash flows. We compare the outcomes of various asset allocation strategies for a typical DC plan investor. The strategies considered include constant proportion, linear glide path, and optimal dynamic (multi-period) time consistent quadratic shortfall approaches. The last of these is based on a double exponential jump diffusion model. We determine the parameters of the model using monthly US data over a 90-year sample period. We carry out tests in a synthetic market which is based on the same jump diffusion model and also using bootstrap resampling of historical data. The probability that portfolio values at retirement will be insufficient to provide adequate retirement incomes is relatively high, unless DC investors adopt optimal allocation strategies and raise typical contribution rates. This suggests there is a looming crisis in DC plans, which requires educating DC plan holders in terms of realistic expectations, required contributions, and optimal asset allocation strategies.

Keywords: defined contribution plan; probability of shortfall; quadratic shortfall; dynamic asset allocation; resampled backtests

JEL Classification: G11

1. Introduction

Defined benefit (DB) pension plans can be high risk liabilities for both private and public sector organizations. To de-risk balance sheets, many employers have been gradually converting to defined contribution (DC) plans. This trend has been particularly strong in Australia and the US: as of 2017, 87% of pension assets in Australia were in DC plans, while the corresponding figure for the US was 60%. In contrast, in countries such as the UK, Canada, and Japan, pension plan assets were still predominantly in DB plans.¹ Of course, considering just the percentage of assets in DC plans at a point in time can hide the underlying trend. For example, about 95% of Canadian pension assets were in DB plans as of 2017. However, Gougeon (2009) points out that the number of DC plan participants in Canada almost doubled between 1991 and 2006, while the number of DB plan participants shrank by about 4% over the same period.

¹ Statistics from the Thinking Ahead Institute's "Global Pension Assets Study 2018". Available online: www.thinkingaheadinstitute.org/-/media/Pdf/TAI/Research-Ideas/GPAS-2018.pdf.

Under a DC plan, the employee and employer contribute to a retirement savings account, which is usually tax-advantaged. In some cases, the employee is able to select from a list of investments (e.g., stock and bond index funds) which have been approved by the employer. However, in terms of the important decisions about asset allocation, the employee is typically left up to her own devices.

Although DC plans are desirable from the employers' point of view, the retirement savings risk has simply been transferred to those who are perhaps least able to manage it. Many studies have confirmed that the average retail investor is very poor at managing investment portfolios (e.g., [Barber and Odean \(2013\)](#)). A typical DC plan investor will accumulate savings for thirty or more years, followed by a decumulation phase of perhaps twenty years. As a result, inefficiencies in asset allocation, even if small in annual terms, can have very large cumulative effects over a typical long-term retirement savings investment horizon.

Target date funds (TDFs) have become a popular means of providing an *autopilot* asset allocation strategy. The prototypical TDF uses a high allocation to equities during the early years of DC savings. The stock allocation is decreased, and the portfolio becomes more weighted towards bonds, as the retirement date nears. The transition from high to low equity allocations is pre-determined by means of a *glide path* strategy. In the US, TDFs are *Qualified Default Investment Alternatives*, implying that the assets of employees enrolled in an employer-managed DC plan can be invested in a TDF by default.² [Holt \(2018\)](#) indicates that total assets in US TDFs at the end of September 2018 were almost \$1.19 trillion, and notes that "given their role as the default investment in many defined contribution retirement plans, investors can expect to see assets in target-date funds continue to rise over the long run".³ This reflects the tendency of many DC plan participants to stick with the default option that is presented to them. This has been noted in the US context by [Madrian and Shea \(2001\)](#) and [Choi et al. \(2002\)](#) amongst several others, and by other authors in various other countries as well.⁴ An implication of investors' choosing default options is that default asset allocations will have a strong effect on the accumulation of savings over time ([Dobrescu et al., 2018](#)). [Choi et al. \(2002\)](#) note that

Sophisticated employers should choose their plan defaults carefully, since these defaults will strongly influence the retirement preparation of their employees. Policymakers should also recognize the role of defaults, since policymakers can facilitate, with laws and regulations, the socially optimal use of defaults ([Choi et al. 2002](#), p. 104).

However, several studies have reported that TDFs do not appear to outperform simpler constant proportion strategies ([Arnott et al., 2013](#); [Basu et al., 2011](#); [Esch and Michaud, 2014](#); [Estrada, 2014](#); [Forsyth and Vetzal, 2019](#)). Clearly, constant proportion strategies which have a significant allocation to equities could result in a wide range of terminal wealth at the end of the accumulation period. The fact that TDFs are not superior to such strategies indicates that TDFs are not a panacea for DC plan investors.

Our objective in this article is to describe realistic outcomes for a typical DC plan investor who contributes a fixed (real) amount per year to an account containing two possible assets, a low risk bond fund and an equity index fund, over a lengthy (30 year) accumulation period, and who follows one of three different asset allocation strategies. The first two of these strategies are often available as default options in US DC plans, these being a constant proportion strategy and a linear glide path, which approximately mimics a prototypical TDF. These strategies are both deterministic since the asset allocation between the bond and equity funds is pre-specified and does not change as a result of realized investment returns. The third strategy is an optimal dynamic (multi-period) quadratic

² In other words, the employee must explicitly decide to opt out of a TDF if she desires a different asset allocation strategy.
³ Although our focus in this article is on the US setting, we note that TDFs are now being marketed in various parts of Europe, in part due to regulatory developments ([Pielichata, 2018](#)). Major US vendors such as Vanguard and Fidelity have launched TDFs in Canada in recent years. In addition, some life-cycle products such as Time Pension which has been popular in Denmark are similar in some respects to TDFs.
⁴ See, for example, [Hedesström et al. \(2004\)](#) (Sweden), [O'Connell \(2009\)](#) (New Zealand), and [Dobrescu et al. \(2018\)](#) (Australia).

shortfall (QS) strategy. While such a strategy is not (to our knowledge) currently available for DC plan investors, it is an interesting extension to consider since it is adaptive, i.e., the asset allocation is a function of the prevailing state of the investment portfolio and the time remaining until anticipated retirement. Our investigation of this strategy allows us to draw conclusions about the extent to which adaptive strategies could potentially improve savings outcomes compared to the types of strategies that are currently popular default options. Unlike the constant proportion and glide path strategies, the QS strategy relies on a parametric model for the equity index. Our implementation models the real (inflation-adjusted) stock index as following a double exponential jump diffusion model (Kou and Wang, 2004), which incorporates the risk of sudden market crashes.⁵ The parameters of the jump diffusion model are estimated by fitting to nine decades of US market data.

The QS strategy is computed by numerically solving a Hamilton–Jacobi–Bellman (HJB) equation (Dang and Forsyth, 2014). It is interesting to note that the QS strategy, which is naturally time consistent, is equivalent to the pre-commitment multi-period mean-variance (MV) strategy, in the sense that both objective functions lead to the same optimal controls. This can be proven based on the embedding technique (Li and Ng, 2000; Zhou and Li, 2000). Our numerical approach allows us to impose realistic no-leverage constraints. This avoids the impractical situation of a highly leveraged portfolio, which often results for unconstrained pre-commitment MV optimal strategies (Lioui and Poncet, 2016)⁶.

As a benchmark case, we consider a constant proportion strategy involving the two assets listed above: a risk-free bond index and a risky stock index. The portfolio is annually rebalanced. The weights on the two assets are chosen to achieve a target expected level of real wealth at the end of the savings period. We then compare this strategy with a glide path and an optimal QS strategy, both of which are constrained to have the same expected value of final wealth as the constant proportion strategy. The glide path and constant proportion asset allocation strategies can be considered to be typical of many DC plan holders. The optimal QS strategy is effectively a *best case* scenario, in terms of reducing the probability of shortfall (over a wide range of the terminal wealth distribution) compared to glide path or constant proportion strategies (Forsyth and Vetzal, 2017b; Forsyth and Vetzal (2019).

We compare these strategies in a *synthetic* market, in which the equity price process is assumed to follow a jump diffusion with constant parameters (i.e., the average historical parameters). We also carry out tests using bootstrap resampling with actual historical data (Cogneau and Zakalmouline, 2013; Dichtl et al., 2016).

We make several assumptions which allow us to convert a desired final salary replacement ratio into a target real portfolio value after a 30-year contribution period. Our base case example assumes a target real final salary replacement ratio of 50%, a combined annual employee-employer contribution of 20% of real salary, and investments in short-term risk-free government bonds and a value-weighted equity index.

Our two main findings are as follows:

1. With typical glide path or constant proportion strategies, there is an unacceptably large probability of shortfall in terms of meeting the target final wealth goal.
2. Even with an optimal dynamic QS asset allocation strategy, there is still a fairly high probability of shortfall. This shortfall probability can be reduced to what we view as a reasonable level by increasing the total contribution rate or reducing the replacement ratio, compared to the base case. Another possibility is to substitute an equal-weighted equity index for the value-weighted index,

⁵ An obvious extension would be a GARCH/stochastic volatility model. However, Ma and Forsyth (2016) document that mean-reverting stochastic volatility effects for Heston-type stochastic volatility models are negligible for long-term investors. Since multivariate GARCH/stochastic volatility models are typically mean-reverting, this suggests that stochastic volatility may be unimportant for long-term investors under these models as well, but this has not been proven.

⁶ Recall the time consistent QS strategy has the same controls as the pre-commitment MV strategy.

but this may not work in practice due to higher costs associated with equal-weighted indexes, which are not recognized in our model.

Our results suggest that most DC plan holders are unlikely to be able to maintain their desired standard of living upon retirement. In our opinion, the true risks of DC plans have not been communicated in realistic terms to plan participants. It seems that this situation can be improved only by simultaneously lowering expectations, increasing savings rates, and improving asset allocation strategies. We emphasize that while, in principle, an optimal QS strategy could be offered as an option to DC plan members by a financial intermediary, and this would give those who selected it improved odds of achieving sufficient levels of retirement savings to maintain their desired living standards in retirement, by itself this is unlikely to be enough. It will also be necessary to raise savings rates and reduce expectations about living standards during retirement.

A possible criticism of our findings is based on the fact that our computations use the past 90 years of US market data. Some would argue that real market returns going forward will be lower than historically observed values. However, this would only lead to worse results, so it could be argued that our conclusions should be even more pessimistic.

2. Formulation

We focus exclusively on a simple context with just two assets available in the financial market, namely a risky asset and a risk-free asset. In practice, the risky asset would be a broad market index fund. An investor saves for retirement at time T . The amounts that this investor’s portfolio contains of the risky and risk-free assets at time t are denoted by S_t and B_t , respectively. The investor’s total wealth from the portfolio (i.e., the total value of the portfolio) at t is $W_t = S_t + B_t$. The fraction of total wealth invested in the risky asset is $p_t = S_t/W_t$. The investment period runs from the inception time $t = 0$ to the horizon date $t = T$. There is a set of $M + 1$ pre-determined *action times* denoted by \mathcal{T} ,

$$\mathcal{T} \equiv \{t_0 = 0 \leq t_1 \leq \dots \leq t_M = T\}. \tag{1}$$

At the horizon date $t_M = T$, the portfolio is liquidated. At each action time $t_i \in \mathcal{T}_1 = \mathcal{T} \setminus \{t_M\}$ (i.e., each action time prior to T), (i) an amount of cash q_i is contributed to the portfolio and then (ii) the portfolio is rebalanced.⁷

Let the instant before action time t_i be $t_i^- = t_i - \epsilon$, where $\epsilon \rightarrow 0^+$. Similarly, the instant after t_i is denoted by $t_i^+ = t_i + \epsilon$. To simplify notation, let $S_i^+ = S_{t_i^+}$, $S_i^- = S_{t_i^-}$, $B_i^+ = B_{t_i^+}$, $B_i^- = B_{t_i^-}$, $W_i^+ = W_{t_i^+}$, and $W_i^- = W_{t_i^-}$. Similarly, let $p_{t_i^+} = p_i$.

Between action times (i.e., $t \notin \mathcal{T}$), the value of the investor’s portfolio will fluctuate in accordance with changes in the unit prices of the two assets. We assume a constant risk-free rate r , so that the evolution of the amount invested in the risk-free asset is

$$dB_t = rB_t dt, \quad t \notin \mathcal{T}. \tag{2}$$

The dynamics of the changes in the amount invested in the risky asset between action times are given by the jump diffusion process

$$\frac{dS_t}{S_t^-} = (\mu - \lambda\kappa) dt + \sigma dZ + d\left(\sum_{i=1}^{\pi_t} (\xi_i - 1)\right), \quad t \notin \mathcal{T}, \tag{3}$$

where μ is the (uncompensated) drift rate, σ is the volatility, dZ is the increment of a Wiener process, π_t is a Poisson process with intensity λ , and ξ denotes the random jump multiplier. When a jump

⁷ As discussed below, in the case of an optimal QS strategy, the investor may also withdraw cash from the portfolio at an action time.

occurs, $S_t = \zeta S_{t-}$, and $\kappa = E[\zeta - 1]$ where $E[\cdot]$ is the expectation operator. We assume that ζ_i are i.i.d. positive random variables characterized by a double exponential distribution (Kou and Wang, 2004). Given that a jump occurs, p_{up} is the probability of an upward jump and $(1 - p_{up})$ is the probability of a downward jump. The density function $f(y = \log \zeta)$ is then

$$f(y) = p_{up}\eta_1 e^{-\eta_1 y} \mathbf{1}_{y \geq 0} + (1 - p_{up})\eta_2 e^{\eta_2 y} \mathbf{1}_{y < 0}. \tag{4}$$

We do not permit short sales of the risky asset, and we impose an upper bound on the use of leverage, i.e., borrowed funds obtained through short sales of the risk-free asset. This means that there is an upper bound on the weight that the investor can place on the risky asset, which we denote by L_{\max} . In other words, $0 \leq p_i \leq L_{\max}$ for all action times $t_i \in \mathcal{T}_1$. Generally, with a DC account, it is reasonable to specify $L_{\max} = 1$, ruling out the use of any leverage. Since the value of the risky asset follows the jump diffusion (3), if we allow leverage by setting $L_{\max} > 1$, the investor can become insolvent.⁸ We add the further constraint that, if the investor becomes insolvent at any time, then trading stops and all positions in the risky asset are liquidated.⁹ In insolvency, debt accumulates until it is (possibly) eliminated by cash contributions. We emphasize that insolvency can only occur if leverage is allowed, i.e., $L_{\max} > 1$.

2.1. Deterministic Glide Paths

TDFs generally use a deterministic glide path, where the asset allocation depends only on time. In our case, this would imply $p_i = p(t)$. One case is a *linear glide path*, with

$$p_i = p_{\max} + \frac{t_i \times (p_{\min} - p_{\max})}{T}, \tag{5}$$

where p_{\max} and p_{\min} are parameters. Note also that a constant proportion strategy can be viewed as a deterministic glide path with $p = \text{const.}$ for all action times $t_i \in \mathcal{T}_1$.

Between action times, the amounts the investor has in the risk-free and risky assets follow the processes (2) and (3), respectively. Recalling that q_i is a cash contribution, at action times prior to the horizon date (i.e., $t_i \in \mathcal{T}_1$), we have

$$\begin{aligned} W_i^+ &= S_i^- + B_i^- + q_i, \\ S_i^+ &= p_i W_i^+, \\ B_i^+ &= (1 - p_i) W_i^+. \end{aligned} \tag{6}$$

In the case of deterministic glide paths, closed form recursive expressions for the mean and variance of terminal wealth W_T are developed in Forsyth and Vetzal (2019). The cumulative distribution function (CDF) for W_T is computed using a Monte Carlo method. In our numerical tests below, we compare all strategies by fixing expected terminal wealth. Since we have closed form expressions for the mean, we determine the glide path parameters using a Newton iteration in order to enforce this condition.¹⁰

⁸ Since the investor rebalances her portfolio discretely, insolvency could also occur if $L_{\max} > 1$ in the special case of the model where jumps are ruled out ($\lambda = 0$), i.e., the value of the risky asset follows geometric Brownian motion.

⁹ More precisely, suppose that insolvency occurs at time t , i.e., $S_t + B_t < 0$. Letting t^+ be the instant after t , then $B_{t^+} = S_t + B_t$ and $S_{t^+} = 0$.

¹⁰ For example, we can exogenously specify p_{\min} and find the value of p_{\max} which generates the desired expected terminal wealth via Newton iteration. Alternatively, we can exogenously set p_{\max} and numerically find the appropriate value of p_{\min} .

2.2. Adaptive Strategies

In contrast to deterministic strategies where the asset allocation depends only on time, adaptive strategies allow the asset allocation to depend on the prevailing state of the investment portfolio. Since we search for the optimal controls over all portfolios with the same wealth after cash injection (W_i^+), this means that $p_i = p_i(W_i^+, t_i^+)$. With an adaptive strategy, it can be optimal to withdraw cash from the portfolio (Cui et al., 2014; Dang and Forsyth, 2016). We denote this optimal cash withdrawal as $c_i \equiv c(W_i^- + q_i, t_i)$. Since we only allow cash withdrawals, $c_i \geq 0$. The control at action time t_i now consists of the pair (p_i, c_i) , i.e., after withdrawing c_i from the portfolio, rebalance so that the fraction invested in the risky asset is p_i .

For explanatory purposes, let us consider first consider a dynamic (multi-period) MV criterion with a specified desired value of $E[W_T] = W_d$. The problem to be solved can be stated as

$$\begin{aligned} \min_{\{(p_0, c_0), \dots, (p_{M-1}, c_{M-1})\}} \quad & \text{Var}[W_T] = E[W_T^2] - W_d^2, \\ \text{subject to} \quad & \begin{cases} \text{At horizon } T: E[W_T = S_T + B_T] = W_d, \\ \text{Between action times } t \notin \mathcal{T}: (B_t, S_t) \text{ follow processes (2), (3),} \\ \text{At action times } t \in \mathcal{T}_1: \\ W_i^+ = S_i^- + B_i^- + q_i - c_i, \\ S_i^+ = p_i W_i^+, \quad B_i^+ = W_i^+ - S_i^+, \\ p_i = p_i(W_i^+, t), \quad 0 \leq p_i \leq L_{\max}, \\ c_i = c_i(W_i^- + q_i, t) \quad c_i \geq 0. \end{cases} \end{aligned} \quad (7)$$

A criticism of the pre-commitment MV problem (7) is that it is *time inconsistent*. In other words, the investor has an incentive to deviate from the strategy computed at time zero (Basak and Chabakauri, 2010). However, in order to solve problem (7), we can use the embedding technique (Li and Ng, 2000; Zhou and Li, 2000). Consider a control set

$$P = \{(p_i(W_i^+, t_i^+), c_i(W_i^- + q_i, t_i)), i = 0, \dots, M - 1\}.$$

Informally, if P^* is an optimal control for problem (7), then there exists a W^* such that P^* is also the optimal control for the following problem:

$$\begin{aligned} \min_{\{(p_0, c_0), \dots, (p_{M-1}, c_{M-1})\}} \quad & E[(W^* - W_T)^2], \\ \text{subject to} \quad & \begin{cases} \text{At horizon } T: E[W_T = S_T + B_T] = W_d, \\ \text{Between action times } t \notin \mathcal{T}: (B_t, S_t) \text{ follow processes (2), (3),} \\ \text{At action times } t \in \mathcal{T}_1: \\ W_i^+ = S_i^- + B_i^- + q_i - c_i, \\ S_i^+ = p_i W_i^+, \quad B_i^+ = W_i^+ - S_i^+, \\ p_i = p_i(W_i^+, t), \quad 0 \leq p_i \leq L_{\max}, \\ c_i = c_i(W_i^- + q_i, t) \quad c_i \geq 0. \end{cases} \end{aligned} \quad (8)$$

Problem (8) can be solved using dynamic programming methods.¹¹

¹¹ If problem (7) is not convex, there may be solutions to problem (8) that are not solutions to problem (7). However, these spurious solutions can be eliminated in a straightforward way (Dang et al., 2016; Tse et al., 2014).

As noted above, it is optimal to withdraw cash from the portfolio under some conditions (Cui et al., 2012; Dang and Forsyth, 2016). Let

$$Q_\ell = \sum_{j=\ell+1}^{j=M-1} e^{-r(t_j-t_\ell)} q_j \tag{9}$$

be the discounted planned future contributions to the DC account at time t_ℓ . If

$$(W_i^- + q_i) > W^* e^{-r(T-t_i)} - Q_i, \tag{10}$$

then the optimal strategy is to

- (i) withdraw cash $c_i = W_i^- + q_i - (W^* e^{-r(T-t_i)} - Q_i)$ from the portfolio; and
- (ii) invest the remainder $(W^* e^{-r(T-t_i)} - Q_i)$ in the risk-free asset.

This is optimal because in this case $E[(W^* - W_T)^2] = 0$, which is the minimum of problem (8).

We refer to any cash withdrawn from the portfolio as *surplus cash* in the following. For the sake of discussion, we will assume that surplus cash is invested in the risk-free asset, but does not contribute to the calculation of the mean and variance of terminal wealth.

Allowing cash withdrawals prevents penalization of wealth paths such that $W_T > W^*$, which can result in forcing the optimal strategy to lose money if market gains are good, which is clearly an undesirable outcome (Cui et al., 2012; Dang and Forsyth, 2016). We remark that in practice, this withdrawal can be virtual, i.e., any amount of wealth satisfying equation (10) is simply invested in the risk-free asset, and the surplus cash is not taken into account when computing the optimal fraction to invest in equities. See Dang and Forsyth (2016) for more detail on this. In fact, if we use continuous rebalancing, then the optimal strategy is such that Equation (10) is never satisfied (Vigna, 2014). In the discrete rebalancing case, the generation of surplus cash is a low probability event.

This target-based approach of problem (8) provides a reasonable objective on its own (Menoncin and Vigna, 2017; Vigna, 2014, 2017). Solving (8) minimizes quadratic shortfall (QS) with respect to W^* , so we will refer to the resulting strategy as the *QS optimal* strategy below. However, this becomes even more compelling when we recall that the solution is also pre-commitment MV efficient. The solution then simultaneously minimizes *two* risk measures: variance around the desired $E[W_T]$ and QS with respect to W^* , as seen at time zero.

We emphasize that the fact that the pre-commitment MV policy is time inconsistent is irrelevant since we take the point of view that we are seeking the QS optimal control, from problem (8). Since the QS problem (8) can be solved using dynamic programming, the controls are trivially time consistent. The fact that the QS problem gives rise to time consistent controls, whereas the MV problem (7) is time inconsistent, is due to the fact that we fix W^* for the QS problem, for all time. At time zero, the MV problem controls and the QS problem controls are the same for W^* computed at time zero. At later times, this correspondence holds only if we allow W^* to change as a function of time. However, using a fixed W^* is intuitively reasonable for DC pension plan saving (Menoncin and Vigna, 2017; Vigna, 2014).

We note that there are techniques for forcing a time consistent constraint for the MV problem (7) (Bjork and Murgoci, 2010, 2014; Bjork et al., 2014; Wang and Forsyth, 2011). However, we prefer the target based QS approach since it is relatively easy to communicate to end user investors (Menoncin and Vigna, 2017; Vigna, 2014). In addition, forcing the time consistent constraint can have result in non-intuitive strategies with strange features (Bensoussan et al., 2019; Wang and Forsyth, 2011).

We formulate problem (8) as the solution of a nonlinear Hamilton–Jacobi–Bellman (HJB) partial integro differential equation. See Dang and Forsyth (2014) for details concerning the numerical solution. Given an arbitrary value of W^* , we can solve problem (8) for the optimal control, which we denote by $P^*(W^*)$. Given the optimal control, cumulative distribution functions are easily found using Monte

Carlo simulation. However, we seek the solution to problem (7), which is expressed in terms of a specified expected value $E[W_T] = W_d$. We determine the value of W^* for problem (8) which satisfies the constraint $E[W_T] = W_d$. We enforce this by a Newton iteration, whereby each function evaluation requires a solution of an HJB equation.

3. Data and Parameter Estimates

The underlying stochastic model outlined above in Equations (2), (3), and (4) involved a constant risk-free rate r for the bond component and a double exponential jump-diffusion for the equity component. Estimation of the parameters of these equations follows the methods described in Forsyth and Vetzal (2019). These procedures are summarized briefly here for convenience. Readers interested in additional details are referred to Forsyth and Vetzal (2019).

We use monthly US data obtained from the Center for Research in Security Prices (CRSP) for the period 1926:1 through 2015:12.¹² Our base case uses the CRSP value-weighted total return index (which includes all distributions for all domestic equities trading on major US exchanges), along with the CRSP 3-month Treasury bill (T-bill) index. The original data are in nominal terms, but we convert them to real terms using the US CPI, also obtained from CRSP. We use real indexes since investors with long-term savings objectives such as funding retirement should concentrate on real (not nominal) wealth goals. For some tests, we use alternative underlying assets: the CRSP equal-weighted total return index (which invests the same amount in each component security, rather than weighting by market capitalization) and a 10-year US Treasury bond (T-bond) index.

Figure 1 provides plots of monthly real returns for the 3-month T-bill, 10-year T-bond, value-weighted total return, and equal-weighted total return indexes. For comparability, all four indexes are plotted with the same vertical axis scale. As expected, the two equity indexes exhibit much higher volatility, with occasional months having returns of large magnitude. This provides a measure of support to our modelling assumptions which assume a constant interest rate but a jump diffusion specification for the equity index. It is also interesting to observe that there was extremely high equity market volatility during the 1930s. By contrast, volatility during the period following the financial crisis that began in 2007 was comparatively mild.

Figure 2 graphs cumulative real returns for the four investment indexes. The vertical axis uses a logarithmic scale. This enhances visibility over time, as otherwise the dramatic growth in the equity indexes over the latter part of the sample obscures the behaviour of those indexes during the earlier periods and renders the behaviour of the two Treasury indexes all but invisible. All four indexes begin at a value of 100 at the start of 1926. The equal-weighted index ends up with the highest value. The historical outperformance of equal-weighting has been attributed to such portfolios having higher exposure to value, size, and market factors (Plyakha et al., 2014). It is also interesting to observe that the 10-year T-bond index had higher cumulative returns than the 3-month T-bill index, but that was entirely due to the post-1980 period: prior to then, the longer maturity T-bond index offered no cumulative advantage over the T-bill index.

Table 1 presents parameter estimates. A threshold method (Cont and Mancini, 2011) was used for the jump diffusion model. These parameter estimates were originally provided in Forsyth and Vetzal (2019), and are reproduced here for convenience. The estimates using the value-weighted equity market index imply an expected real annual return of close to 9%, about 3% lower than the corresponding value for the equal-weighted index. Of course, the price to be paid for this difference is higher risk. The equal-weighted index shows higher diffusive volatility (σ). Since jumps are expected to occur on average every $1/\lambda$ years, the equal-weighted index tends to have jumps a bit more often. Conditional

¹² More precisely, our calculations are based on data from Historical Indexes, ©2015 Center for Research in Security Prices (CRSP), The University of Chicago Booth School of Business. Wharton Research Data Services (WRDS) was used in preparing this work. This service and the data available thereon constitute valuable intellectual property and trade secrets of WRDS and/or its third-party suppliers.

on a jump occurring, it is much more likely to be a downward jump in each case. Average jump magnitudes are $1/\eta_1$ for upward jumps and $1/\eta_2$ for downward jumps, and these are both larger for the equal-weighted index. A similar comment applies to the standard deviation of the jump size since this is equal to the mean for the exponential distribution. Turning to the bond market indexes, Table 1 shows that the long run average real annual return for the 10-year T-bond index was just over 2%, while that for the shorter maturity index was around 80 basis points. Of course, these higher returns are accompanied by higher volatility, as indicated by the top two plots of Figure 1.

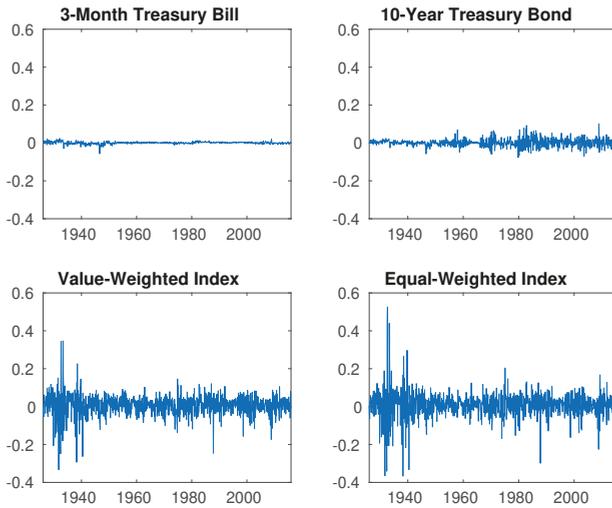


Figure 1. Monthly real returns for US investment indexes, 1926–2015.

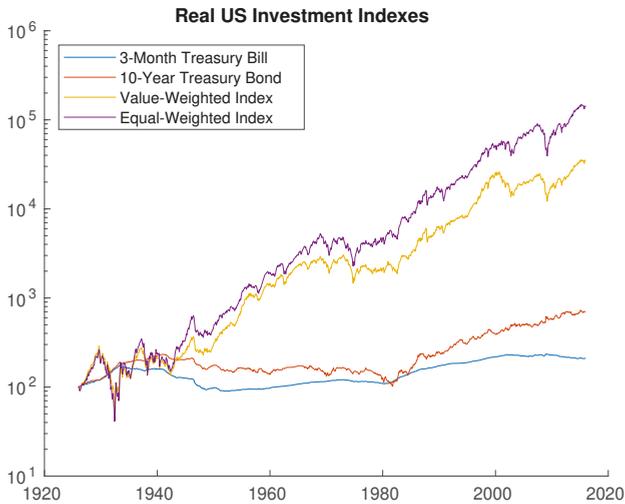


Figure 2. Cumulative real returns for US investment indexes, 1926–2015.

Table 1. Annualized parameter estimates based on real monthly data from 1926:1 to 2015:12. These values originally appeared in Forsyth and Vetzal (2019) and are reproduced here for convenience. Parameters for the equity market indexes were estimated using the threshold technique of Cont and Mancini (2011). The average returns for the bond indexes were calculated as $\log[B(T)/B(0)]/T$, where $B(t)$ denotes the index level at the time t .

Equity Market Index	μ	σ	λ	p_{up}	η_1	η_2
Value-weighted	0.0889	0.1477	0.3222	0.2759	4.4273	5.2613
Equal-weighted	0.1183	0.1663	0.4000	0.3333	3.6912	4.5409
Bond Market Index		Average Return				
3-month T-bill		0.00827				
10-year T-bond		0.02160				

3.1. Robustness to Parameter Estimation

Our main purpose for calibrating the parameters for the stochastic processes (3) is to determine a control strategy (i.e., fraction in risky asset at rebalancing dates). Consequently, our concern is with the effect of calibration errors on the computed strategy, rather than minimizing fit errors in an econometric sense. In Forsyth and Vetzal (2017a) and in Dang and Forsyth (2016), an extensive study of the effect of parameter ambiguity is carried out. In particular, the parameters for stochastic process (3) were determined using maximum likelihood, and threshold techniques with various parameters. Robustness of the strategy (in the synthetic market) was tested by using Monte Carlo simulations with different parameters than were used in computing the optimal strategy. For example, the strategy was computed and stored, assuming parameters determined from the threshold strategy. Then, this strategy was tested using Monte Carlo simulations, in a synthetic market driven by a stochastic process with parameters determined using maximum likelihood. In other words, the parameters used to compute the strategy were misspecified. In all combinations of methods, the results were robust to this type of parameter misspecification.

However, the real test of our strategies is their performance on bootstrapped resampled tests. In the bootstrap tests, we make no assumptions about the stochastic process followed in the historical market. However, in all our bootstrap tests, the adaptive quadratic shortfall strategy (computed using the estimated market parameters) outperforms the glide path and constant proportion strategies.

4. Base Case Scenario

We consider an example where a DC plan member wants to generate retirement income (in real terms) of a specified fraction R of final salary. Studies have shown that earnings for a typical employee increase rapidly until the age of 35, then increase slowly thereafter, until a few years before retirement, and then decrease as fewer hours are worked in the transition to retirement (Cocco et al., 2005; Ruppert and Zanella, 2015). Given an initial real annual salary I_0 , we assume that real annual income in year t before retirement is given by $I_t = e^{\mu t} I_0$. In other words, salary is expected to grow in real terms at a constant annual rate μ_I . Upon retirement at time T , real annual salary is $e^{\mu_I T} I_0$. Our assumptions would be relevant to a 35 year old employee with stable employment, who intends to work full-time until the age of 65.

To determine the amount of real wealth required to fund this replacement income during retirement, we use the well-known 4% rule of Bengen (1994). Bengen examined historical data to determine the maximum real withdrawal rate that a retiree could safely use without exhausting her assets over a 35 year period. Bengen assumed that accumulated pension wealth was invested in a portfolio having half invested in stocks and half invested in intermediate-term US Treasury securities, and concluded that a 4% withdrawal rate (escalated by the rate of inflation) was quite safe.

Dang et al. (2017) recently revisited this rule. The problem was posed somewhat differently: the idea was to determine a real withdrawal rate such that half of the real wealth at the start of

retirement remained after 20 years, with high probability. Rather than using fixed portfolio weights as in Bengen (1994), the portfolio was invested in stocks and bonds according to the QS optimal strategy described above. Dang et al. concluded that the 4% rule still held up well under the revised assumptions.

An obvious alternative way to generate retirement cash flows is to buy a lifetime annuity. However, in practice most retirees are not willing to do this, for a variety of reasons (MacDonald et al., 2013). In the current environment of low real interest rates, annuities provide rather low income, so the reluctance of retirees to use them is particularly unsurprising.

Consequently, we pose the pension accumulation problem as follows. The desired expected accumulated real wealth at retirement W_d is

$$E[W_T] = W_d = \frac{R e^{\mu_I T} I_0}{w_r}, \tag{11}$$

where R is the replacement ratio, I_0 is the initial salary, μ_I is the real salary escalation rate, T is the end of the accumulation period, and w_r is the safe withdrawal rate. Recall that above we denoted the set of action times prior to T as \mathcal{T}_1 . We assume that cash is contributed into the portfolio at time $t_i \in \mathcal{T}_1$. The initial cash contribution at t_0 is $F_c I_0$, where F_c is the real contribution fraction. This contribution fraction represents the total amount contributed by both the employee and the employer to all retirement savings accounts, but excludes any government sponsored universal schemes (e.g., CPP in Canada, Social Security in the US). We also assume that these accounts are tax-advantaged, i.e., no tax is paid during the accumulation phase. At subsequent action times $t_i \in \mathcal{T}_1$ (i.e., after the initial contribution), the amount contributed is assumed to be $F_c I_0 e^{\mu_I t_i}$.

Table 2 summarizes the data for this base case scenario. As indicated in the table, we assume a replacement fraction of 50%, an initial salary of \$50,000, and a real escalation rate of 1.27%.¹³ Combined contributions by the employee and employer to the retirement savings portfolio are 20% of real salary each year. For simplicity, these contributions are assumed to be made at the start of each year during a 30 year accumulation period. The safe withdrawal rate is assumed to be 4%, in line with Bengen (1994) and Dang et al. (2017). Applying these parameters to Equation (11), we find that the expected (real) desired terminal wealth W_d is

$$E[W_T] = W_d = \frac{0.50}{0.04} \times \$50,000 e^{0.0127 \times 30} \simeq \$915,000. \tag{12}$$

Note that the specification given in Equation (11) implies that decreasing the withdrawal rate w_r has the same effect as increasing the replacement fraction R or the salary escalation rate μ_I .

Table 2. Data for base case scenario. Cash is injected into the portfolio at times $t = 0, 1, \dots, 29$. Market parameters for the equity and bond indexes are provided in Table 1.

Initial salary I_0	\$50,000
Salary escalation rate μ_I	0.0127 (Bloom et al., 2014)
Contribution fraction F_c	0.20
Accumulation period T	30 years
Safe withdrawal rate w_r	0.04
Equity index	Value-weighted
Bond index	3-month T-bill
Investment strategies	Constant proportion, glide path, QS optimal
Rebalancing interval	1 year
Maximum leverage indicator L_{\max}	1.0
If insolvent	Trading stops

¹³ As noted by Bloom et al. (2014), this rate has been used by the US Congressional Budget Office in its long-term projections.

As indicated in Table 2, the retirement savings portfolio is invested in the value-weighted index and the 3-month T-bill index. Relevant parameters for these indexes are given in Table 1. We consider three alternative investment strategies:

- Constant proportion, i.e., $p = const$.
- Linear glide path, as in Equation (5).
- Time consistent QS optimal strategy, as described in Section 2.2. Recall that this strategy is also multi-period pre-commitment MV optimal.

Each of these strategies is rebalanced annually. We start with the constant proportion strategy, and determine the equity weight such that $E[W_T] = \$915,000$, assuming the market parameters given in Table 1 for the value-weighted equity index and the 3-month T-bill index. This turns out to be 0.5788.¹⁴ We then turn to the linear glide path strategy. In this case, we specify $p_{max} = 1.0$, and then determine that the value of $p_{min} = 0.3066$ is needed to have $E[W_T] = \$915,000$. We proceed similarly for the QS optimal strategy given by the solution to problem (8). Imposing the leverage constraint $L_{max} = 1.0$, we find by Newton iteration that the value of W^* which results in $E[W_T] = \$915,000$ is \$1,106,200. We compute and store the optimal control associated with this value of W^* in order to test this strategy. Note that:

- When we change input parameters (e.g., invest in different assets, allow $L_{max} > 1$, etc.), we may need to recompute the expected wealth target $E[W_T] = W_d$, the equity weight for the constant proportion strategy, the glide path parameters (p_{max}, p_{min}) , and the quadratic wealth target W^* (along with the associated optimal control) in order to meet this target.
- The quadratic wealth target W^* exceeds the target expected real terminal wealth $W_d = E[W_T]$. This is because the QS optimal strategy will de-risk if W^* is attainable by investing only in the risk-free asset, so there is not much chance of exceeding this quadratic target by a significant amount. This implies that the average terminal wealth, factoring in paths where the accumulated savings does not ever reach W^* , must be lower than W^* .

We compare these three strategies using two different types of simulation tests. As an initial test, we assume that the stochastic environment described in Table 1 holds exactly. In other words, the level of the equity market index follows a double exponential jump diffusion with the parameters given in Table 1 and the bond market index is non-stochastic, with a constant risk-free interest rate as indicated in that table. We refer to this as a *synthetic market*. In such a market, we draw 160,000 Monte Carlo simulated paths and compute performance statistics. Note that these comparisons are based on a simulated environment that corresponds exactly to the environment used to formulate the strategies. As a second and more stringent test, we draw simulated paths by bootstrap resampling of the historical return data and compute the same performance statistics. We refer to this type of backtest as a *historical market*. This is a stricter test since it does not assume that the equity market follows a jump diffusion process or that the risk-free interest rate is constant over time, although those assumptions are still used to generate the strategies that are followed. A single resampled path is constructed by pasting together enough blocks of monthly historical return data to cover the investment horizon of 30 years. The sampling is done in blocks to account for possible serial dependence. The blocks are selected simultaneously from both the historical stock and bond market indexes, to incorporate possible correlations. The blocks are chosen randomly, with replacement. To avoid end issues, the historical data is wrapped around.¹⁵ To reduce the impact of a fixed blocksize and mitigate edge effects at each block end, we use the stationary block bootstrap (Patton et al., 2009;

¹⁴ This is a bit more aggressive in terms of taking on equity market risk than the strategy considered by Bengen (1994) which involved equal weights between the equity and bond markets. Keep in mind that here we are investing in a 3-month T-bill index, whereas Bengen used intermediate maturity Treasury bonds which offer somewhat higher average returns.

¹⁵ In other words, if the size of a block extends past the end of the sample in 2015:12, the return data resumes at the start of the sample in 1926:1 for the duration of the block.

Politis and White, 2004) where the blocksize is sampled randomly from a geometric distribution with an expected value of \hat{b} . In principle, the optimal expected blocksize can be estimated using an algorithm provided by Patton et al. (2009). As discussed in Forsyth and Vetzal (2019), this approach is not easily applied in our context. This is because the estimated optimal blocksizes for the different market indexes we consider (i.e., the value-weighted and equal-weighted US equity market indexes and the 3-month T-bill and 10-year T-bond indexes) vary considerably, ranging from about two months for the value-weighted stock index to more than four years for the T-bill index. Recall that we sample simultaneously from both a stock index and a bond index, so we must use the same blocksize for both indexes, and our strategies involve weighted combinations of two of these indexes that can change deterministically (glide path) and also randomly (QS optimal) along a simulation path. As a result, we report results for a range of expected blocksizes \hat{b} , acknowledging that the choice of \hat{b} for our application is open to debate.

Table 3 gives the results for the base case input data from Table 2. Consider first the synthetic market. By construction, all three strategies have the same expected value of real terminal wealth at the retirement date of 915 (units in the table are in thousands of dollars). The constant proportion and glide path strategies are effectively indistinguishable, having the same standard deviation of terminal wealth and the same probability of ending up with wealth below 700 or below 800. By contrast, the QS optimal strategy has much lower standard deviation or shortfall probability for those two levels of terminal wealth, by a factor of around two in each case. In addition, this strategy offers a small amount of expected surplus cash (this surplus is not applicable for the other two strategies). Turning to the historical market results for expected blocksizes of 1, 2, and 5 years, we reach generally similar conclusions. Of course, the expected values no longer are exactly equal to the target of 915, but the difference from this target is much lower for the QS optimal strategy. The constant proportion and glide path strategies are again quite comparable to each other, with the glide path having slightly lower expected value and standard deviation, but a bit higher shortfall probability (for the two values of W_T considered). In all cases, the QS optimal strategy offers the best performance, with higher mean, lower standard deviation, and lower shortfall probability, as well as a modest amount of expected surplus cash.

Table 3. Base case scenario results. Wealth units: thousands of dollars. Input data provided in Tables 1 and 2. Synthetic market results computed using Monte Carlo simulations with 160,000 sample paths. Historical market results based on 10,000 bootstrap resampled paths using data from 1926:1 to 2015:12.

Strategy	Expected Value	Standard Deviation	$Pr(W_T < 700)$	$Pr(W_T < 800)$	Expected Surplus Cash
Synthetic Market					
Constant proportion	915	519	0.39	0.51	NA
Glide path	915	519	0.39	0.51	NA
QS optimal	915	244	0.19	0.24	21
Historical Market (Expected blocksize $\hat{b} = 1$ year)					
Constant proportion	876	402	0.38	0.51	NA
Glide path	872	398	0.39	0.52	NA
QS optimal	904	232	0.18	0.24	26
Historical Market (Expected blocksize $\hat{b} = 2$ years)					
Constant proportion	869	376	0.38	0.51	NA
Glide path	866	372	0.39	0.52	NA
QS optimal	911	221	0.17	0.23	31
Historical Market (Expected blocksize $\hat{b} = 5$ years)					
Constant proportion	862	349	0.37	0.50	NA
Glide path	861	347	0.38	0.51	NA
QS optimal	924	213	0.16	0.21	38

Figure 3a shows the cumulative distributions for all three strategies in the historical market (expected blocksize of two years). The distributions for the constant proportion and glide path strategies are practically identical. Over the bulk of the distribution, the QS optimal strategy exhibits clearly better performance than the others. However, the QS optimal strategy performs a little worse than the alternatives in the extreme left tail.¹⁶ This happens because there can be paths where the equity market trends downward for a very large portion of the investment horizon. In such cases, all strategies do poorly, but the QS optimal strategy will remain fully invested in the equity market in an attempt to recover and meet the quadratic wealth target. The QS optimal strategy also underperforms in the extreme right tail of the distribution. This is because there are paths where the equity market trends strongly upward over most of the investment period. Once the quadratic wealth target is reached, however, the QS optimal strategy de-risks, shifting all investment into the low return bond market. It does not capitalize on the continued strong equity market performance. The other two strategies, by comparison, retain a large equity market exposure, leading to higher terminal wealth. However, we reiterate that, over most of the distribution, the QS optimal strategy provides better results.

Figure 3b depicts properties of the optimal control for the QS optimal strategy. This strategy invests entirely in the equity market for the first several years. The percentage invested in the risky asset subsequently trends downward on average over time. However, there is considerable variation: the standard deviation of the optimal control rises strongly over time, indicating that the allocation to equities is quite sensitive to realized investment returns.¹⁷

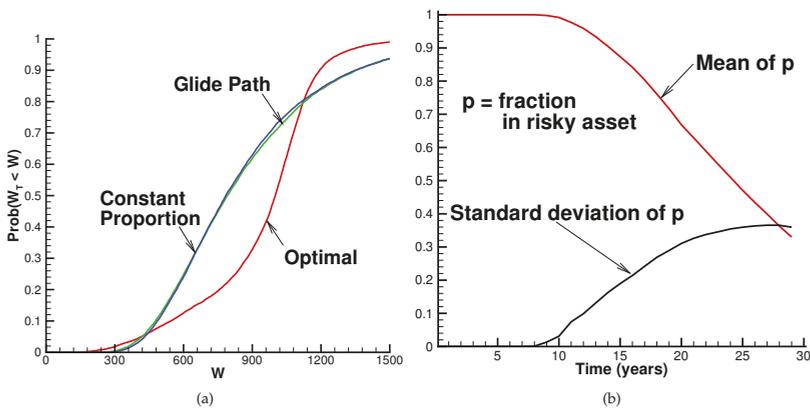


Figure 3. Base case scenario results in the historical market. Input data provided in Tables 1 and 2. Results based on 10,000 bootstrap resampled paths using data from 1926:1 to 2015:12 with expected blocksize $\hat{b} = 2$ years. (a) cumulative distributions of real terminal wealth for various strategies. Wealth units: thousands of dollars; surplus cash included for the QS optimal case; (b) mean and standard deviation of the fraction allocated to the equity market for the QS optimal strategy.

The historical market results given in Table 3 are somewhat encouraging for the QS optimal strategy. To formulate this strategy, recall that we assumed a double exponential jump diffusion model with known parameters for the equity market index and a constant risk-free interest rate. Our bootstrap resampled historical market tests make no such assumptions, yet deliver results that are fairly close to

¹⁶ In other words, the QS optimal strategy will appear riskier than the constant proportion or glide path strategies according to tail risk measures such as value-at-risk or conditional value-at-risk, provided that the risk measure is calculated using sufficiently low cumulative probabilities.

¹⁷ Of course, the equity allocation for the constant proportion and glide path cases is fixed in advance, being at most time-dependent and not varying at all in response to realized returns.

those observed in the synthetic market tests which maintain these assumptions. This indicates that the QS optimal strategy is quite robust to departures from these assumptions.

As an additional test of the robustness of the QS optimal strategy, we explore the effect of computing and storing the optimal control based on constant parameters from Table 1 but then allowing the synthetic equity market parameters to vary randomly in simulation tests. To be specific, we carry out Monte Carlo simulations where at each action time $t_i \in \mathcal{T}_1$ and along each stochastic path we select (μ, σ) from a uniform distribution having mean equal to the corresponding values from Table 1. This set of (μ, σ) is then used for the interval (t_i, t_{i+1}) .

Table 4 shows the results. The first row reproduces the values reported in Table 3 for this strategy in the synthetic market, while the remaining two rows provide results when μ and σ are varied randomly within the given range. Table 4 indicates that only an estimate of the mean of the distribution of the market parameters μ and σ is needed to compute an effective control strategy. This is consistent with the results in [Ma and Forsyth \(2016\)](#), where it is shown that including stochastic volatility effects results in negligible improvements in results for long-term (i.e., greater than ten years) investors.

Table 4. Base case scenario results for the QS optimal strategy with random variation of market parameters μ and σ , Wealth units: thousands of dollars. Input data provided in Tables 1 and 2, except as noted. Monte Carlo simulations with (μ, σ) drawn from a uniform distribution with the indicated limits along each of 160,000 paths.

Market Parameters	Expected Value	Standard Deviation	$Pr(W_T < 700)$	$Pr(W_T < 800)$	Expected Surplus Cash
Synthetic market	915	244	0.19	0.24	21
$\mu \in [0.04889, 0.01289]$	916	245	0.19	0.24	21
$\sigma \in [0.1077, 0.1877]$	915	245	0.19	0.24	21

Overall, then, for the base case data in Table 2, the QS optimal strategy appears to be fairly robust to parameter and model mis-specification. Moreover, this strategy clearly outperforms the constant proportion and glide path alternatives over most of the terminal wealth distribution. It provides about the same mean terminal wealth, but has considerably lower standard deviation and shortfall probability (for the two wealth levels considered in Table 3). However, even the QS optimal strategy delivers somewhat disappointing results in absolute terms. Recall that we are trying to achieve average real terminal wealth of \$915,000. In the idealized synthetic market which conforms exactly to the modelling assumptions used to generate the controls for the QS optimal strategy, Table 3 shows that there is almost a 20% chance of ending up with real terminal wealth below \$700,000, and about a 25% chance of ending up with less than \$800,000. These shortfall probabilities do not change much under the conditions of the historical market backtests. Of course, these rather pessimistic results so far have only considered the base case data. In the following section, we investigate whether more promising results can be achieved under different assumptions.

5. Alternative Assumptions

The base case results in Section 4 relied on the input parameters given in Tables 1 and 2, for the value-weighted equity and 3-month T-bill indexes. We now explore the effects of altering our assumptions about factors such as the contribution fraction, the salary escalation rate, the maximum amount of leverage permitted, the underlying indexes to be used, and the salary replacement ratio. We consider each of these in turn. For the most part, we only use the QS optimal strategy since it has been shown above to be generally superior to the constant proportion and glide path strategies.

5.1. Effect of Contribution Fraction

Our base case described by Table 2 assumed a total combined contribution by the employee and employer of $F_c = 20\%$ of salary. Table 5 reports the effects of dropping this to 15% or increasing it to 25%

for the QS optimal strategy. As is to be expected, the table shows that risk (measured either in terms of standard deviation or the reported shortfall probabilities) decreases significantly as F_c rises. However, even in the case where 25% of the employee’s salary is contributed to the retirement savings plan, there is still almost a 15% chance that real terminal wealth is less than \$800,000, considerably lower than the target of \$915,000. A broader comparison of these cases is provided in Figure 4, which depicts the cumulative distributions of real terminal wealth. The cases where $F_c = 20\%$ and $F_c = 25\%$ appear quite comparable for values of $W_T \geq \$915,000$ over the plotted range, but the higher contribution fraction case appears to be much safer over a wide range below the target. The high savings rate leads to a notably increased amount of expected surplus cash. Along paths with strong equity market returns, the target can be reached relatively early, but the model assumes that savings continue each year (in this case at a high rate), so surplus cash can build up. The case where $F_c = 15\%$ exhibits poor performance on the downside, but is somewhat better on the upside. With a low amount of money saved, more risk must be taken on in order to reach the target. Doing so works out very well if realized returns are strong, and quite poorly if they are not.

Table 5. Effect of varying contribution fraction F_c for the QS optimal strategy. Wealth units: thousands of dollars. Input data provided in Tables 1 and 2, except as noted. Synthetic market results computed using Monte Carlo simulations with 160,000 sample paths. Historical market results based on 10,000 bootstrap resampled paths using data from 1926:1 to 2015:12.

F_c	Expected Value	Standard Deviation	$Pr(W_T < 700)$	$Pr(W_T < 800)$	Expected Surplus Cash
Synthetic Market					
0.15	915	440	0.36	0.42	18
0.20	915	244	0.19	0.24	21
0.25	915	150	0.09	0.13	21
Historical Market (Expected blocksize $\hat{b} = 2$ years)					
0.15	916	380	0.34	0.41	12
0.20	911	221	0.17	0.23	31
0.25	909	126	0.07	0.14	67

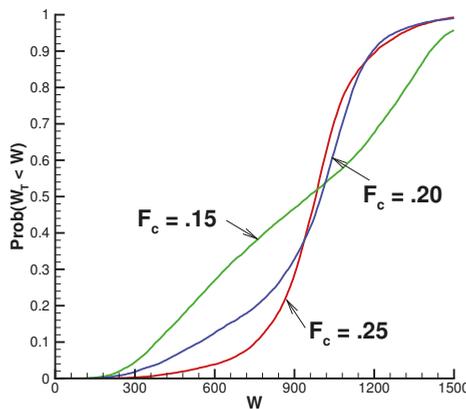


Figure 4. Cumulative distributions of real terminal wealth for various contribution fractions F_c for the QS optimal strategy. Wealth units: thousands of dollars. Input data provided in Tables 1 and 2, except as noted. Historical market results based on 10,000 bootstrap resampled paths using data from 1926:1 to 2015:12 with expected blocksize $\hat{b} = 2$ years; surplus cash included.

5.2. *Effect of Salary Escalation Rate*

We now examine the effect of changing the salary escalation rate μ_I from the base case value of 1.27% given in Table 2. Note that this will result in a different expected real wealth target W_d , based on Equation (11). Table 6 summarizes the results for both the synthetic market and the historical market, using the QS optimal control. For comparability, the table expresses the results in terms of W_T/W_d , i.e., real terminal wealth as a fraction of the expected wealth target. Obviously, a higher escalation rate leads to a higher final salary. Given a fixed salary replacement ratio, this translates into a higher expected value of terminal wealth W_d . For example, with $\mu_I = 1.75\%$, we have $W_d = \$1,056,000$ instead of $\$915,000$ as in the base case. The results in Table 6 are quite similar in the synthetic and historical markets, on a case by case basis. For each set of tests, the standard deviation and shortfall probabilities show an increase with μ_I . With a higher salary and a fixed contribution fraction, there will obviously be a higher amount of saving. Despite this, the associated higher real terminal wealth target results in a higher level of risk. This is borne out in Figure 5, which plots the cumulative distribution functions of normalized real terminal wealth W_T/W_d for the various values of μ_I . The highest salary escalation rate has the worst performance for low W_T/W_d , and the best performance for high W_T/W_d . Taking on more risk to reach the higher wealth target works out well if investment returns are favourable, and poorly if they are not. Conversely, the lowest value of μ_I results in the best performance if investment returns are weak, and the worst performance if they are not.

Table 6. Effect of varying salary escalation rate μ_I for the QS optimal strategy. Units for W_d : thousands of dollars. Remaining wealth values are normalized by W_d for each case. Input data provided in Tables 1 and 2, except as noted. Synthetic market results computed using Monte Carlo simulations with 160,000 sample paths. Historical market results based on 10,000 bootstrap resampled paths using data from 1926:1 to 2015:12.

μ_I	Wealth Target W_d	Expected Value (W_T/W_d)	Standard Deviation (W_T/W_d)	$Pr(W_T/W_d < 0.8)$	$Pr(W_T/W_d < 0.9)$	Expected Surplus Cash $(/W_d)$
Synthetic Market						
0.0175	1056	1.0	0.31	0.25	0.30	0.02
0.0127	915	1.0	0.27	0.20	0.25	0.02
0.0075	783	1.0	0.22	0.15	0.21	0.02
Historical Market (Expected blocksize $\hat{b} = 2$ years)						
0.0175	1056	0.99	0.29	0.23	0.29	0.03
0.0127	915	0.97	0.24	0.18	0.25	0.03
0.0075	783	0.99	0.19	0.14	0.21	0.04

5.3. *Effect of Leverage*

All of the results presented thus far have specified $L_{max} = 1$, as indicated in Table 2. We now consider increasing this to $L_{max} = 1.5$, thereby allowing the use of 50% leverage. Table 7 documents the results for the QS optimal strategy. It is interesting to observe that the risk measures shown (standard deviation of W_T and the two shortfall probabilities) each indicate somewhat lower risk when the use of leverage is permitted. In one sense, this is obvious: relaxing a constraint cannot lead to worse performance. On the other hand, leverage is often not allowed since it is perceived to be “risky”. The issue here is what is meant by risk. We have defined it on the basis of the value of terminal wealth. Leverage constraints, however, are typically motivated by concerns that portfolio values during the accumulation period (not at the end of the period) will fluctuate excessively.

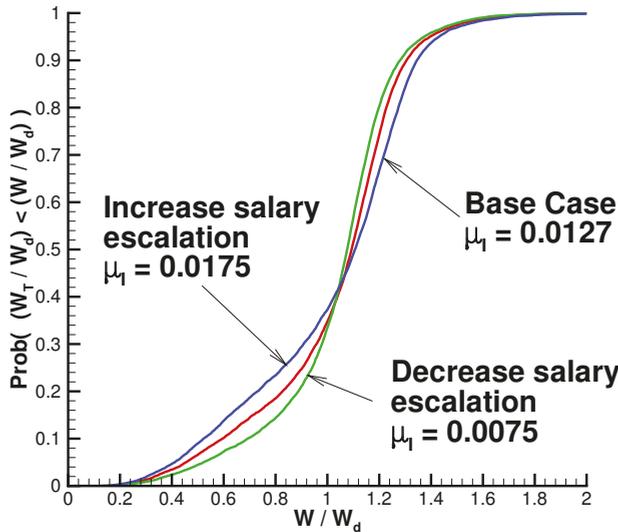


Figure 5. Cumulative distributions of normalized real terminal wealth for various salary escalation rates μ_1 for the QS optimal strategy. Input data provided in Tables 1 and 2, except as noted. Historical market results based on 10,000 bootstrap resampled paths using data from 1926:1 to 2015:12 with expected blocksize $\hat{b} = 2$ years; surplus cash included.

Table 7. Effect of varying maximum leverage indicator L_{max} for the QS optimal strategy. Wealth units: thousands of dollars. Input data provided in Tables 1 and 2, except as noted. Synthetic market results computed using Monte Carlo simulations with 160,000 sample paths. Historical market results based on 10,000 bootstrap resampled paths using data from 1926:1 to 2015:12.

L_{max}	Expected Value	Standard Deviation	$Pr(W_T < 700)$	$Pr(W_T < 800)$	Expected Surplus Cash
Synthetic Market					
1.0	915	244	0.19	0.24	21
1.5	915	205	0.12	0.17	24
Historical Market (Expected blocksize $\hat{b} = 2$ years)					
1.0	911	221	0.17	0.23	31
1.5	904	186	0.11	0.18	46

Figure 6a shows the cumulative distributions of real terminal wealth when leverage is allowed and when it is not. The cumulative distributions are fairly similar for high levels of terminal wealth. Over a wide range of wealth levels below the target of \$915,000, the strategy which permits leverage performs better. However, in the extreme left tail of the distribution, it turns out to be worse to allow leverage. These very low wealth levels occur as a result of very poor equity market returns over most of the investment horizon. The QS optimal strategy continues to try to reach the quadratic wealth target, so it invests completely in the equity market to the extent possible. With continued poor returns, leverage in this case amounts to doubling down on a losing bet.

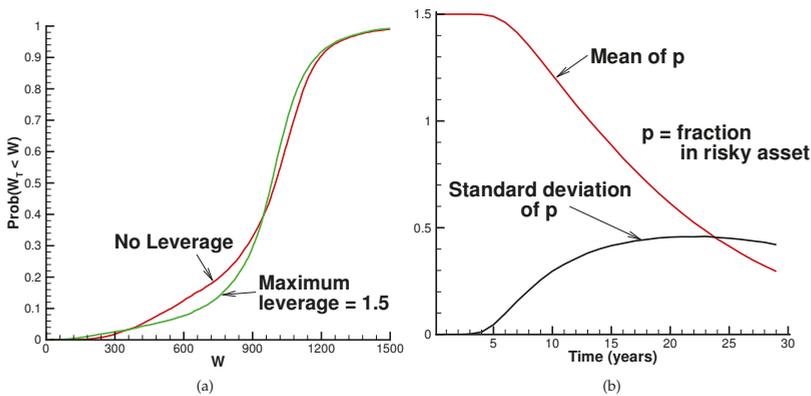


Figure 6. Results for cases allowing and excluding leverage. Input data provided in Tables 1 and 2, except as noted. Historical market results based on 10,000 bootstrap resampled paths using data from 1926:1 to 2015:12 with expected blocksize $\hat{b} = 2$ years. (a) cumulative distribution function of real terminal wealth for cases allowing leverage ($L_{\max} = 1.5$) and excluding it ($L_{\max} = 1$). Wealth units: thousands of dollars; surplus cash included; (b) mean and standard deviation of the fraction allocated to the equity market for the case where leverage ($L_{\max} = 1.5$) is allowed.

5.4. Long-Term Bond Index

All of the results presented thus far were based on the value-weighted equity index and the 3-month T-bill index. In this section, we explore the effects of substituting the 10-year T-bond index for the T-bill index. Recall from Table 1 that the long-term bond index had average annual real returns of 2.16%, about 1.3% higher than the T-bill index. Of course, these higher average returns are accompanied by higher volatility (Figure 1), but that is ignored when we formulate the QS optimal strategy since we assume constant interest rates. Table 8 shows results for the constant proportion, linear glide path, and QS optimal strategies when the 10-year T-bond index is used in lieu of the 3-month T-bill index. The general pattern in Table 8 is similar to that seen earlier in Table 3. The QS optimal approach achieves roughly the same expected real terminal wealth, but takes on considerably lower risk compared to the other two strategies, as measured by standard deviation of W_T or either of the two shortfall probabilities considered. Comparing the results from Tables 3 and 8, we observe that the standard deviation and shortfall probabilities for each strategy are a little lower in the synthetic market when the 10-year T-bond index is used. This also holds in the historical market for the constant proportion and glide path strategies, but is not the case for the QS optimal strategy in terms of the shortfall probabilities. For example, the probability of real terminal wealth being lower than \$800,000 is 23% if the 3-month T-bill index is used, but 26% when the 10-year T-bond index is used. The overall conclusion, however, is that replacing the 3-month T-bill index by the 10-year T-bond index does not make much difference. This picture is reinforced by comparing the cumulative distributions of real terminal wealth shown in Figure 7 based on the 10-year T-bond index with those shown previously in Figure 3a for the 3-month T-bill index.

Table 8. Results for the optimal QS strategy when the 10-year T-bond index is used instead of the 3-month T-bill index. Wealth units: thousands of dollars. Input data provided in Tables 1 and 2, except as noted. Synthetic market results computed using Monte Carlo simulations with 160,000 sample paths. Historical market results based on 10,000 bootstrap resampled paths using data from 1926:1 to 2015:12.

Strategy	Expected Value	Standard Deviation	$Pr(W_T < 700)$	$Pr(W_T < 800)$	Expected Surplus Cash
Synthetic Market					
Constant proportion	915	437	0.34	0.47	NA
Glide path	915	438	0.34	0.48	NA
QS optimal	915	222	0.16	0.21	19
Historical Market (Expected blocksize $\hat{b} = 2$ years)					
Constant proportion	900	374	0.34	0.47	NA
Glide path	897	376	0.34	0.48	NA
QS optimal	881	201	0.17	0.26	67

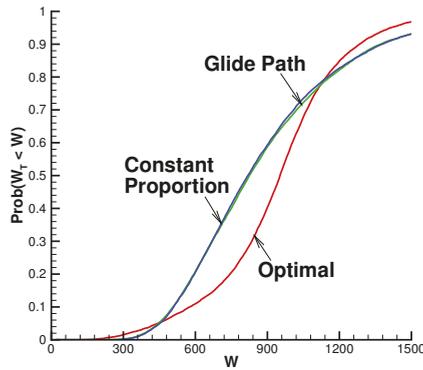


Figure 7. Cumulative distribution of real terminal wealth for QS optimal, glide path, and constant proportion strategies when the 10-year T-bond index is used. Wealth units: thousands of dollars. Input data provided in Tables 1 and 2, except as noted; surplus cash included for the QS optimal strategy. Historical market results based on 10,000 bootstrap resampled paths using data from 1926:1 to 2015:12. Expected blocksize $\hat{b} = 2$ years.

5.5. Equal-Weighted Equity Index

As shown above in Figure 2, an equal-weighted equity index has historically outperformed the value-weighted equity index that has been used for all results to this point. We now consider the impact of replacing the value-weighted index by its equal-weighted counterpart. Note that the bond index used here is the 3-month T-bill index. Table 9 provides the results. Whether in the idealized synthetic market or the backtest historical market, the QS optimal strategy clearly outperforms the constant proportion and glide path alternatives by the criteria given in the table. In the synthetic market, the QS optimal strategy achieves the same $E[W_T]$ with dramatically lower standard deviation of W_T and shortfall probability, along with the possibility of a modest amount of surplus cash. The same general conclusions apply in the historical market, although it is worth noting that the average real terminal wealth for the other strategies is somewhat lower for all expected blocksizes considered. Comparing the results in Table 9 for the equal-weighted equity index with those reported above in Table 3 for the value-weighted index, it can be seen that the shortfall probabilities are now considerably lower for the QS optimal strategy, but almost unchanged for the other strategies. The standard deviation of W_T , however, is substantially lower for all of the strategies.

Table 9. Results when the equal-weighted equity index is used instead of the value-weighted equity index. Wealth units: thousands of dollars. Input data provided in Tables 1 and 2, except as noted. Synthetic market results computed using Monte Carlo simulations with 160,000 sample paths. Historical market results based on 10,000 bootstrap resampled paths using data from 1926:1 to 2015:12.

Strategy	Expected Value	Standard Deviation	$Pr(W_T < 700)$	$Pr(W_T < 800)$	Expected Surplus Cash
Synthetic Market					
Constant proportion	915	546	0.38	0.51	NA
Glide path	915	553	0.39	0.52	NA
QS optimal	915	185	0.11	0.16	44
Historical Market (Expected blocksize $\hat{b} = 1$ year)					
Constant proportion	837	327	0.39	0.54	NA
Glide path	831	319	0.39	0.55	NA
QS optimal	904	162	0.10	0.16	49
Historical Market (Expected blocksize $\hat{b} = 2$ years)					
Constant proportion	827	293	0.38	.54	NA
Glide path	820	283	0.38	0.55	NA
QS optimal	915	139	0.07	0.13	51
Historical Market (Expected blocksize $\hat{b} = 5$ years)					
Constant proportion	815	248	0.36	0.53	NA
Glide path	808	242	0.37	0.55	NA
QS optimal	932	115	0.04	0.10	54

Figure 8 plots the cumulative distributions of real terminal wealth in both the synthetic and the historical markets. In both cases, the distributions for the glide path and constant proportion strategies are virtually indistinguishable. Figure 8a indicates that the QS optimal strategy outperforms over a wide range of terminal wealth values in the synthetic market, although it does perform worse in the tails of the distribution. As mentioned earlier, this is due to two features of the strategy: (i) it automatically de-risks once the quadratic wealth target is achievable by investing only in the bond index (so it does not take advantage of continued strong equity market performance afterwards on paths where that happens); and (ii) it continually tries to reach the quadratic wealth target by using maximum equity market exposure (and this gamble for resurrection fails on paths where the equity market has persistently poor performance). The same comments apply to the historical market shown in Figure 8b, but it is worth noting that the underperformance of the QS optimal strategy in the tails is considerably reduced here compared to the synthetic market.

5.6. Effect of Replacement Ratio

All of the results provided to here assume a replacement ratio $R = 50\%$ of final real salary, in accordance with Table 2. We now explore the effects of lowering this to 40% and increasing it to 60%. For each case, we determine the desired expected real wealth target by using Equation (11). Table 10 shows the results. Decreasing R to 40% reduces the expected wealth target W_d to \$732,000 from \$915,000, while raising R to 60% increases W_d to \$1,098,000. The remaining wealth values in the table are normalized by W_d . Whether we consider the synthetic or the historical market, it is clear that increasing R requires taking on more risk, as measured by either the standard deviation or the shortfall probabilities. This is borne out in the cumulative distribution plots of normalized terminal real wealth provided in Figure 9. The synthetic market results in Figure 9a clearly indicate that the QS optimal strategy performs better for the lowest value of R . Recall that the strategy attempts to come as close as possible to W_d (normalized to 1 for this plot), and the cumulative distribution when $R = 40\%$ shows relatively low probability of normalized real wealth being much above or below 1. For the highest replacement ratio ($R = 60\%$), there is a substantial chance of being either significantly below or above 1. This is because the strategy must take on more risk in order to attain the higher expected wealth

target. Of course, the base case with $R = 50\%$ lies in between these other two cases. The results for the historical market shown in Figure 9b are generally similar, though the differences across the range of values of R are somewhat less pronounced.

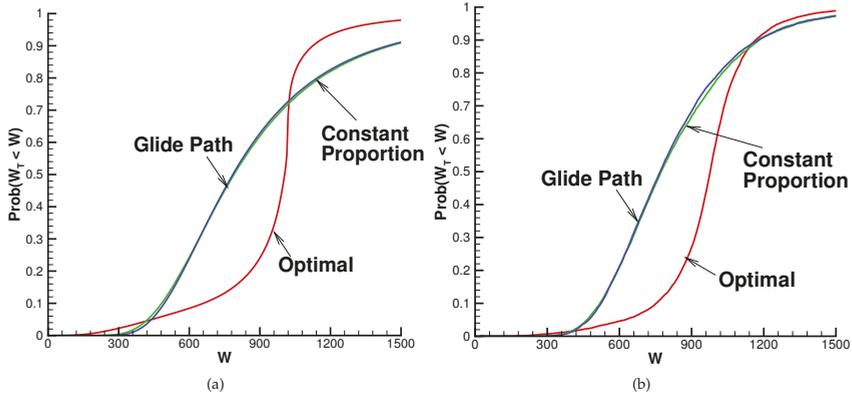


Figure 8. Cumulative distribution of real terminal wealth when the equal-weighted equity index is used. Wealth units: thousands of dollars. Input data provided in Tables 1 and 2, except as noted. Synthetic market results computed using Monte Carlo simulations with 160,000 sample paths. Historical market results based on 10,000 bootstrap resampled paths using data from 1926:1 to 2015:12 with expected blocksize $\hat{b} = 2$ years; surplus cash flow included for the QS optimal strategy. (a) synthetic market; (b) historical market.

Table 10. QS optimal results with varying salary replacement ratios R . Units for W_d : thousands of dollars. Remaining wealth values are normalized by W_d for each case. Input data provided in Tables 1 and 2, except as noted. Synthetic market results computed using Monte Carlo simulations with 160,000 sample paths. Historical market results based on 10,000 bootstrap resampled paths using data from 1926:1 to 2015:12.

R	Wealth Target W_d	Expected Value (W_T/W_d)	Standard Deviation (W_T/W_d)	$Pr(W_T/W_d < 0.8)$	$Pr(W_T/W_d < 0.9)$	Expected Surplus Cash $(/W_d)$
Synthetic Market						
0.4	732	1.0	0.16	0.10	0.15	0.02
0.5	915	1.0	0.27	0.20	0.25	0.02
0.6	1098	1.0	0.37	0.31	0.36	0.02
Historical Market (Expected blocksize $\hat{b} = 2$ years)						
0.4	732	.99	0.15	0.09	0.17	0.0
0.5	915	.97	0.24	0.18	0.25	0.03
0.6	1098	1.0	0.35	0.30	0.35	0.02

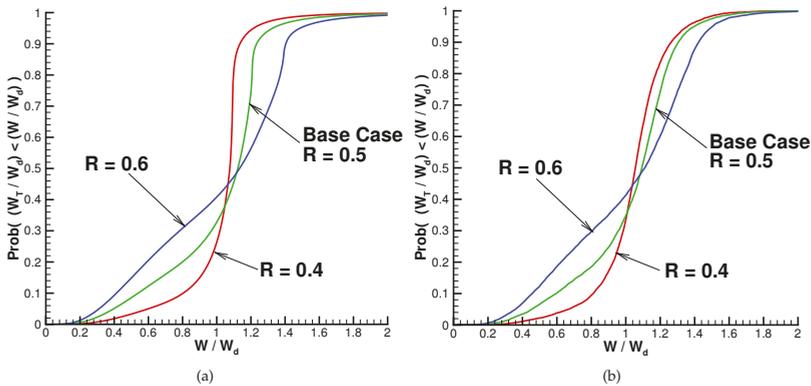


Figure 9. Cumulative distributions of real terminal wealth with different salary replacement ratios R for the QS optimal strategy. Input data provided in Tables 1 and 2, except as noted. Synthetic market results computed using Monte Carlo simulations with 160,000 sample paths. Historical market results based on 10,000 bootstrap resampled paths using historical data from 1926:1 to 2015:12 with expected blocksize $\hat{b} = 2$ years; surplus cash included. (a) synthetic market; b) historical market.

5.7. Summary Regarding Alternative Assumptions

Sections 5.1–5.6 above provided detailed results concerning the effects of

- varying the accumulation fraction F_c , i.e., the investor saves $F_c I_0 e^{\mu t_i}$ at each rebalancing date;
- varying the real salary escalation rate μ_I ;
- use of leverage;
- alternative bond index: use of a 10 year T-bond index instead of a 3 month T-bill index;
- alternative stock index: use of an equal-weighted equity index instead of a value-weighted index; and
- varying the replacement ratio R .

Table 11 summarizes the results of various strategies in terms of probability of shortfall with respect to the desired wealth goal W_d . These results were all obtained using bootstrap resampling (i.e., for the historical market) with an expected blocksize of two years.

It is interesting to note that the results for the Base Case, constant proportion strategy (contribution rate 20%) are worse than the results for the QS optimal strategy, $F_c = 0.15$ (contribution rate 15%), at least in terms of the two points of the cumulative distribution function listed in the table. In other words, the shortfall increase with a constant proportion strategy compared to the quadratic shortfall strategy can be interpreted as losing 5% of lifetime salary, which is very significant. However, this comparison does not take into account the entire cumulative distribution function. In general, constant proportion strategies are superior to quadratic shortfall policies in the extreme left tail of the distribution. However, the improvement over quadratic shortfall is very small, with a very low probability.

As a filter to determine an acceptable combination of DC plan parameters and investment strategies, suppose we specify that there should be at least a 90% probability of achieving at least 80% of the desired expected wealth goal W_d . Based on attempting to achieve the final target expected real wealth for the base case (see Table 2) and applying this filter, we can see that the shortfall probabilities using standard strategies (constant proportion or glide path) are unacceptably high. Using the QS optimal strategy leads to a substantial reduction in these shortfall probabilities, but still not to the desirable range of less than 10%.

From Equation (11), it is clear that the case with $R = 0.6$ and $w_r = 4\%$ leads to the same expected wealth target as specifying $R = 0.5$ and $w_r = 3.3\%$. Table 11 therefore indicates that, if we assume that the safe (real) withdrawal rate is 3.3% and the replacement ratio is 50%, the probability of shortfall is quite high even if the QS optimal strategy is followed.

Table 11. Comparison of shortfall probabilities. Results are normalized by W_d for each case. Input data provided in Tables 1 and 2, except as noted. Historical market results based on 10,000 bootstrap resampled paths using data from 1926:1 to 2015:12 with expected blocksize $\hat{b} = 2$ years.

Case	$Pr(W_T/W_d < 0.8)$	$Pr(W_T/W_d < 0.9)$
Base Case, Table 2		
Constant proportion	0.43	0.54
Glide path	0.43	0.54
QS optimal	0.18	0.25
QS Optimal		
Contribution Fraction		
$F_c = 0.25$	0.09	0.17
$F_c = 0.15$	0.37	0.43
Salary Escalation Rate		
$\mu_I = 0.0175$	0.23	0.29
$\mu_I = 0.0075$	0.14	0.21
Leverage		
$L_{max} = 1.5$	0.13	0.19
Alternative Bond Index		
10-year T-bond	0.19	0.28
Alternative Stock Index		
Equal-weighted	0.09	0.16
Replacement Ratio		
$R = 0.4$	0.09	0.17
$R = 0.6$	0.30	0.35

Assuming we use the QS optimal allocation strategy, we are then forced to take other actions to attempt to reduce the shortfall probability. Increasing the contribution rate to 25% of annual salary meets our criterion, but this might be difficult to implement in terms of agreement from employees and employers. Decreasing the replacement ratio (40% of final salary) also achieves the shortfall objective. We note that many institutions effectively do this by targeting a final career average salary replacement ratio (instead of a final salary replacement ratio).

Finally, the use of the alternative equal-weighted equity index also achieves the shortfall probability target. As noted earlier, this type of index has historically outperformed its value-weighted counterpart owing to higher exposure to value, size, and market factors (Plyakha et al., 2014). In effect, the equal-weighted portfolio is a *smart beta* portfolio, with a long track record. However, equal-weighted portfolios have higher costs, which have not been factored in to our analysis. This suggests that there may be a market opportunity for a low cost synthetic ETF which tracks the equal-weighted index.

6. Conclusions

Typical default strategies available to DC plan members include target date funds (i.e., glide path) and balanced funds (constant proportion). Both of these strategies have high probabilities of shortfall and large standard deviations of real terminal wealth.

We define an acceptable probability of success as a 90% probability of achieving 80% of the real desired terminal wealth goal. QS optimal strategies reduce the probability of shortfall significantly, compared to the ubiquitous glide path and balanced portfolio strategies, but not to within an acceptable range.

Assuming an optimal QS strategy, acceptable probabilities of shortfall can be obtained by

- reducing the final salary target replacement ratio (40% or less);
- increasing the total (employee and employer) contribution rate to 25% per year;
- using alternative stock investment indices, such as an equal-weighted index. The backtests of an equal-weighted index perform well, but it is not clear that this will persist in the future. In addition, we have not factored in the additional costs of this type of index.

Our main conclusion is that current practices in DC plans with typical target benefits and default investment strategies have unacceptably high probabilities of shortfall. This has major ramifications for the many organizations which are shifting to DC plans.

Our results are based on the past 90 years of US market data. It is possible to argue that future market returns will be lower than observed in the past. The implication of such an assumption is that the probability of success for a DC investor would be even lower than we have reported here. In this case, the situation for DC plan investors would be even more dire than we suggest.

Author Contributions: All authors contributed equally in all aspects to this paper.

Funding: P. A. Forsyth's work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) grant RGPIN-2017-03760.

Conflicts of Interest: The authors have no conflicts of interest to report.

References

- Arnott, Robert D., Katrina F. Sherrerd, and Lilian Wu. 2013. The glidepath illusion and potential solutions. *The Journal of Retirement* 1: 13–28. [\[CrossRef\]](#)
- Barber, Brad M., and Terrance Odean. 2013. The behavior of individual investors. In *Handbook of Economics and Finance*. Edited by George Constantinides, Milton Harris and Rene Stulz. Amsterdam: Elsevier, chp. 22, pp. 1533–69.
- Basak, Suleyman, and Georgy Chabakauri. 2010. Dynamic mean-variance asset allocation. *Review of Financial Studies* 23: 2970–3016. [\[CrossRef\]](#)
- Basu, Anup K., Alistair Byrne, and Michael E. Drew. 2011. Dynamic lifecycle strategies for target date retirement funds. *Journal of Portfolio Management* 37: 83–96. [\[CrossRef\]](#)
- Bengen, William. 1994. Determining withdrawal rates using historical data. *Journal of Financial Planning* 7: 171–180.
- Bensoussan, Alain, Kwok Chuen Wong, and Sheung Chi Phillip Yam. 2019. A paradox in time-consistency in the mean-variance problem? *Finance and Stochastics* 23: 173–207. [\[CrossRef\]](#)
- Björk, Tomas, and Agatha Murgoci. 2010. *A General Theory of Markovian Time Inconsistent Stochastic Control Problems*. SSRN 1694759. Rochester: Social Science Research Network.
- Björk, Tomas, and Agatha Murgoci. 2014. A theory of Markovian time inconsistent stochastic control in discrete time. *Finance and Stochastics* 18: 545–92. [\[CrossRef\]](#)
- Björk, Tomas, Agatha Murgoci, and Xun Yu Zhou. 2014. Mean-variance portfolio optimization with state-dependent risk aversion. *Mathematical Finance* 24: 1–24. [\[CrossRef\]](#)
- Bloom, David E., David Canning, and Michael Moore. 2014. Optimal retirement with increasing longevity. *Scandinavian Journal of Economics* 116: 838–58. [\[CrossRef\]](#)
- Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick. 2002. Defined contribution pensions: Plan rules, participant choices, and the path of least resistance. *Tax Policy and the Economy* 16: 67–113. [\[CrossRef\]](#)
- Cocco, João F., Francisco J. Goems, and Pascal J. Maenhout. 2005. Consumption and portfolio choice over the life cycle. *Review of Financial Studies* 18: 491–533. [\[CrossRef\]](#)
- Cogneau, Phillipe, and Valereiy Zakamouline. 2013. Block bootstrap methods and the choice of stocks for the long run. *Quantitative Finance* 13: 1443–57. [\[CrossRef\]](#)
- Cont, Rama, and Cecilia Mancini. 2011. Nonparametric tests for pathwise properties of semimartingales. *Bernoulli* 17: 781–813. [\[CrossRef\]](#)
- Cui, Xiangyu, Jianjun Gao, Xun Li, and Duan Li. 2014. Optimal multi-period mean variance policy under no-shorting constraint. *European Journal of Operational Research* 234: 459–68. [\[CrossRef\]](#)
- Cui, Xiangyu, Duan Li, Shouyang Wang, and Shushang Zhu. 2012. Better than dynamic mean-variance: Time inconsistency and free cash flow stream. *Mathematical Finance* 22: 346–78. [\[CrossRef\]](#)
- Dang, Duy-Minh, and Peter A. Forsyth. 2016. Better than pre-commitment mean-variance portfolio allocation strategies: A semi-self-financing Hamilton–Jacobi–Bellman equation approach. *European Journal of Operational Research* 250: 827–41. [\[CrossRef\]](#)
- Dang, Duy-Minh, Peter A. Forsyth, and Kenneth R. Vetzal. 2017. The 4% strategy revisited: A pre-commitment optimal mean-variance approach to wealth management data. *Quantitative Finance* 17: 335–51. [\[CrossRef\]](#)

- Dang, Duy-Minh, and Peter A. Forsyth. 2014. Continuous time mean-variance optimal portfolio allocation under jump diffusion: A numerical impulse control approach. *Numerical Methods for Partial Differential Equations* 30: 664–98. [CrossRef]
- Dang, Duy-Minh, Peter A. Forsyth, and Yuying Li. 2016. Convergence of the embedded mean-variance optimal points with discrete sampling. *Numerische Mathematik* 132: 271–302. [CrossRef]
- Dichtl, Hubert, Wolfgang Drobetz, and Martin Wambach. 2016. Testing rebalancing strategies for stock-bond portfolios across different asset allocations. *Applied Economics* 48: 772–88. [CrossRef]
- Dobrescu, Loretta I., Xiaodong Fan, Hazel Bateman, Ben R. Newell, A. Ortmann, and Susan Thorp. 2018. Retirement savings: A tale of decisions and defaults. *Economic Journal* 128: 1047–94. [CrossRef]
- Esch, David N., and Robert O. Michaud. 2014. *The False Promise Of Target Date Funds*. Working Paper. Boston: New Frontier Advisors, LLC.
- Estrada, Javier. 2014. The glidepath illusion: An international perspective. *Journal of Portfolio Management* 40: 52–64. [CrossRef]
- Forsyth, Peter A., and Kenneth R. Vetzal. 2017a. Dynamic mean variance asset allocation: Tests for robustness. *International Journal of Financial Engineering* 4: 1750021. [CrossRef]
- Forsyth, Peter A., and Kenneth R. Vetzal. 2017b. Robust asset allocation for long-term target-based investing. *International Journal of Theoretical and Applied Finance* 20: 1750017. [CrossRef]
- Forsyth, Peter A., and Kenneth R. Vetzal. 2019. Optimal asset allocation for retirement saving: Deterministic vs. time consistent adaptive strategies. *Applied Mathematical Finance*, forthcoming. [CrossRef]
- Gougeon, Phillipe. 2009. Shifting pensions. *Statistics Canada: Perspectives on Labour and Income* 10: 16–23.
- Hedesström, Ted M., Henrik Svedsåter, and Tommy Gärling. 2004. Identifying heuristic choice rules in the Swedish premium pension scheme. *Journal of Behavioral Finance* 5: 32–42. [CrossRef]
- Holt, Jeff. 2018. Target-Date Funds Keep Growing In Popularity. Available online: www.morningstar.com/videos/889903/targetdate-funds-keep-growing-in-popularity.html (accessed 9 April 2019).
- Kou, Steven, and Hui Wang. 2004. Option pricing under a double exponential jump diffusion model. *Management Science* 50: 1178–92. [CrossRef]
- Li, Duan, and Wan-Lung Ng. 2000. Optimal dynamic portfolio selection: Multiperiod mean-variance formulation. *Mathematical Finance* 10: 387–406. [CrossRef]
- Lioui, Abraham, and Patrice Poncet. 2016. Understanding dynamic mean variance asset allocation. *European Journal of Operational Research* 254: 320–37. [CrossRef]
- Ma, Kai, and Peter A. Forsyth. 2016. Numerical solution of the Hamilton–Jacobi–Bellman formulation for continuous time mean variance asset allocation under stochastic volatility. *Journal of Computational Finance* 20: 1–37. [CrossRef]
- MacDonald, Bonnie-Jean, Bruce Jones, Richaard J. Morrison, Robert L. Brown, and Mary Hardy. 2013. Research and reality: A literature review on drawing down retirement savings. *North American Actuarial Journal* 17: 181–215. [CrossRef]
- Madrian, Brigitte C., and Dennis F. Shea. 2001. The power of suggestion: Inertia in 401(k) participation and savings behavior. *Quarterly Journal of Economics* 116: 1149–87. [CrossRef]
- Menoncin, Francesco, and Elena Vigna. 2017. Mean-variance target based optimisation for defined contribution pension schemes in a stochastic framework. *Insurance: Mathematics and Economics* 76: 172–84. [CrossRef]
- O’Connell, Allison. 2009. Kiwisaver: A model scheme? *Social Policy Journal of New Zealand* 36: 130–41.
- Patton, Andrew, Dimitris Politis, and Halbert White. 2009. Correction to: Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* 28: 372–375. [CrossRef]
- Pielichata, Paulina. 2018. Europeans seeing funds as solution to DC dilemma. *Pensions & Investments*, March 19.
- Plyakha, Yuliya, Raman Uppal, and Grigory Vilkov. 2014. *Equal or Value Weighting? Implications For Asset-Pricing Tests*. Working Paper. Lille, Nice and Paris: EDHEC Business School.
- Politis, Dimitris, and Halbert White. 2004. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews* 23: 53–70. [CrossRef]
- Ruppert, Peter, and Giulio Zanella. 2015. Revisiting wage, earnings, and hours profiles. *Journal of Monetary Economics* 72: 114–30. [CrossRef]
- Tse, Shu-Tong, Peter A. Forsyth, and Yuying Li. 2014. Preservation of scalarization optimal points in the embedding technique for continuous time mean variance optimization. *SIAM Journal on Control and Optimization* 52: 1527–46. [CrossRef]

- Vigna, Elena. 2014. On efficiency of mean-variance based portfolio selection in defined contribution pension schemes. *Quantitative Finance* 14: 237–58. [[CrossRef](#)]
- Vigna, Elena. 2017. *Tail Optimality And Preferences Consistency For Intertemporal Optimization Problems*. Working Paper No. 502. Collegio Carlo Alberto: Università Degli Studi di Torino.
- Wang, Jian, and Peter A. Forsyth. 2011. Continuous time mean variance asset allocation: A time-consistent strategy. *European Journal of Operational Research* 209: 184–201. [[CrossRef](#)]
- Zhou, Xun Yu, and Duan Li. 2000. Continuous-time mean-variance portfolio selection: A stochastic LQ framework. *Applied Mathematics and Optimization* 42: 19–33. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Article

The Impact of Algorithmic Trading in a Simulated Asset Market

Purba Mukerji ^{1,*}, Christine Chung ², Timothy Walsh ² and Bo Xiong ²

¹ Department of Economics, Connecticut College, New London, CT 06320, USA

² Department of Computer Science, Connecticut College, New London, CT 06320, USA;
cchung@conncoll.edu (C.C.); twalsh@alumni.conncoll.edu (T.W.); bxiong@alumni.conncoll.edu (B.X.)

* Correspondence: pmukerji@conncoll.edu

Received: 13 March 2019; Accepted: 17 April 2019; Published: 20 April 2019

Abstract: In this work we simulate algorithmic trading (AT) in asset markets to clarify its impact. Our markets consist of human and algorithmic counterparts of traders that trade based on technical and fundamental analysis, and statistical arbitrage strategies. Our specific contributions are: (1) directly analyze AT behavior to connect AT trading strategies to specific outcomes in the market; (2) measure the impact of AT on market quality; and (3) test the sensitivity of our findings to variations in market conditions and possible future events of interest. Examples of such variations and future events are the level of market uncertainty and the degree of algorithmic versus human trading. Our results show that liquidity increases initially as AT rises to about 10% share of the market; beyond this point, liquidity increases only marginally. Statistical arbitrage appears to lead to significant deviation from fundamentals. Our results can facilitate market oversight and provide hypotheses for future empirical work charting the path for developing countries where AT is still at a nascent stage.

Keywords: algorithmic trading; market quality; liquidity; statistical arbitrage

1. Introduction

Algorithmic trading (AT) in US stock markets has grown at a blistering pace, starting from the mid-1990s when it accounted for only 3 percent of the market, to recent times when it has reached almost 85% of dollar trade volumes (Zhang 2010). What we learn from markets like the US where AT has progressed very far, can be practically useful in charting the path ahead for developing countries, where AT is still at a nascent stage (Sherry 2017).

The aim of this paper is to bridge the gap between researchers' empirical results and the real-world attempts (e.g., by regulators) to understand how each algorithmic trading strategy works. We do this by using simulations to systematically demonstrate how the behavior of various algorithms result in the observed impact on market quality (liquidity, price spreads, trade-related price discovery, and correlation of asset prices).

Specifically, our simulated markets give us the flexibility to start from the AT strategies of our choosing and simulate the resulting market outcomes. In this way, our work attempts to evaluate if AT strategies can indeed lead to the market outcomes that empirical studies have deduced from market data. Our line of research can also aid market regulators by providing a platform capable of isolating and testing the market impact of specific AT strategies.

Empirical analysis of algorithmic trading have estimated its impact on various aspects of market quality (see for example, Brogaard 2010; Brogaard et al. 2014; Chaboud et al. 2014; Hasbrouck and Saar 2013; Hendershott et al. 2011; Hendershott and Riordan 2011; Kelejian and Mukerji 2016; Menkveld 2013; Riordan and Storkenmaier 2011; Zhang 2010). However, many of the findings are contradictory. These contradictions in the empirical literature may be the result of differences in methodologies, time periods, and samples of assets analyzed. Even in the same work, there are often differences in the degree to which

the impact of algorithmic trading is felt in among different types of assets, and transactions. In addition, for each empirical finding there are different explanations that could produce the observed impact of algorithmic trading.

In its actual work in the practical world, algorithmic trading has made rapid progress in technology, and this has led to an arms-race among participants for acquiring the fastest and most efficient algorithms and machines (Hasbrouck and Saar 2013). As a side effect, increased competition has eroded profits. Regulators have also clamped down on algorithmic trading, following accusations of market manipulation (McCrank 2015). Market manipulation, once thought of as a predominantly developing market issue (Sohel Azad et al. 2014), has now featured quite prominently in developed countries via AT. One of the outcomes of this scenario is that AT is resorting to high risk strategies in hopes of making profits (Philips 2013).

These factors indicate that it would be useful to go a step beyond measuring AT's impact and attempt to unravel how this impact is actually created by the operation of different strategies in AT. Toward this ultimate goal, in this work, we create a computer simulation of the asset market. Our computer simulation gives us the capability to characterize the impact of AT on market quality and test its sensitivity to changing situations such as the volume of algorithmic versus human trading. Since proprietary trading algorithms are not public information, we make simplifying assumptions about the traders that allow us to approximate their behavior. Naturally, the algorithms that actually operate in the marketplace have developed nuances that this exercise does not capture. However, we expect that our results still provide a useful estimation of overarching and longer-term market outcomes.

We model two trader types: fundamental analysts and technical analysts. Both types of traders can be either human or algorithmic. We simulate the behavior of traders to test the liquidity and price discovery in the market as the percentage of algorithmic traders' increases. As a measure of liquidity, we use effective half-spread (e.g., Hendershott and Riordan 2011). Our results indicate that liquidity improves as the proportion of algorithmic traders increase, although most of the liquidity increase is achieved at very low levels of AT. We also investigate statistical arbitrage, which is a commonly used AT strategy, where algorithmic traders use short-term price correlations to predict price movements, and trade to profit from them. Our simulations measure the extent to which this may lead to price movements unrelated to fundamentals. Our findings indicate a significant movement away from fundamentals, with the rise of AT.

Thus, our findings match the real world where AT has changed the landscape of market transactions and has had profound implications for market participants (e.g., Kelejian and Mukerji 2016), and for those charged with market oversight (e.g., U.S. Commodities Futures Trading Commission and U.S. Securities and Exchange Commission 2010). The reason for the changes being experienced is the important differences that set human and algorithmic traders apart, even though their strategies might seem similar. First, AT can make decisions based on much more information and much faster than human traders (Hasbrouck and Saar 2013). Second, human traders use their judgement in addition to trading strategies. AT, on the other hand, makes decisions mechanically. This mechanical decision making may sometimes lead to absurd and undesirable outcomes (Zweig 2010). Recognizing this shortcoming in AT, firms designed safeguards where, under defined conditions, there is a trigger that makes the AT stop for human input (U.S. Commodities Futures Trading Commission and U.S. Securities and Exchange Commission 2010). However, the system is far from perfect and has led to occasional turmoil in the markets. For example, investigation into the "flash crash" of 2010 revealed that a set of algorithmic trades had led to a cascading of prices and extreme volatility. This triggered a practical shut down of markets as ATs paused for human input (U.S. Commodities Futures Trading Commission and U.S. Securities and Exchange Commission 2010).

By providing a clear focus on how AT strategies lead to market outcomes, our findings deepen our understanding of the real-world workings of the asset market, as the participation of AT increases. They serve to clarify existing empirical results in the literature, as well as to suggest future avenues

of empirical research. Specifically, our findings in both experiments provide some support for the existing empirical results in the literature. In each case, they also provide practical nuances in the results, which can be used to form new hypotheses for future empirical work.

The rest of the paper is laid out as follows. Section 2 presents the literature review. Section 3 lays out our simulated market model, while Section 4 presents the experiments based on varying the degree of AT participation and evaluating the impact of AT strategies. Section 5 concludes the paper.

2. Related Literature

The literature in the area of algorithmic trading is vast, relative to its recent vintage (for example, see (Biais and Woolley 2011), for background and literature survey). In this section we briefly discuss some studies that are most closely related to ours.

In the finance literature there are a number of studies into market quality resulting from AT. Here are some that help to illustrate the diversity of data used and the existence of sometimes contradictory results. Hendershott and Riordan (2011) using 13 trading days data in January 2008 on AT orders submitted in the Deutsche Boerse, and Hendershott et al. (2011) using NYSE (New York Stock Exchange) electronic message traffic data over 2001–2005 concluded that as AT grows, liquidity improves. They also found that the initial price impact of an algorithmic trade is larger than that of human orders. Brogaard (2010) analyzed data from the financial crisis in 2008–2009 and concluded that AT reduces volatility and contributes significantly to the price discovery process. Kelejian and Mukerji (2016) analyzed the long-term impact of AT on non-AT traders with data from before the advent of AT to recent times (1985–2012) and found that AT transmits volatility based on short-term price correlations, moving the market away from fundamentals. In a study of minute-by-minute exchange rate trading data over the two years 2006–2007, Chaboud et al. (2014) concluded that non algorithmic traders cause most of the long run variance in the price of currencies, thus proving themselves to be better informed than algorithmic traders. They also found that AT is associated with lower liquidity immediately following important macroeconomic data releases. The research by Foucault and Menkveld (2008) is an example of a study of how upgrades to the trading system lead to different impacts on the market outcomes such as liquidity.

As stated in the introduction, our paper's contribution is in seeking to bridge the gap between the empirical studies such as those we just outlined above and the actual working of AT. As our simulations of AT strategies reveal the associated market outcomes, we provide a cross-check for the conclusions that empirical studies have drawn from market data.

Our work also relates to asset market simulation studies; for example, Kearns et al. (2010) studied the profitability of AT. Using stock market data from 2008, they simulated an "omniscient" AT trader that effectively overestimates the maximum possible profit that could be earned from AT traders that year. They concluded that the total profit derived from AT trading is in general quite modest relative to the total trading volume. Arthur et al. (1996) created a stock market simulation environment called the Santa Fe Artificial Stock Market, which is an example of an agent-based financial market. Autonomous agents in this market learn and adapt their strategies over time. The focus is on the performance of the agents and their strategies and not on the market outcomes.

In contrast to these existing works that simulate asset markets, our novel contribution is that we study the impact of AT on market quality, while the existing literature is geared towards working out the profitability and performance of AT and market agents who use it. Thus, our line of research can ultimately aid regulatory interventions into market manipulations, by providing a platform capable of isolating and testing the market impact of specific strategies.

3. Experimental Method

Real data, from traders and trading outcomes, is used at every stage of our analysis. Each trader's strategy is based on real world strategies employed by actual traders. The market composition of algorithmic versus human traders is set up to simulate the growth of actual algorithmic trading over

time, which researchers in the field have quantified. Finally, each experiment’s findings have been carefully compared to real data based empirical findings from the literature.

In this section we present the settings of our simulation environment: the trader types and parameter settings for each, the strategies that each of our traders uses, and the rules for deciding the volume of shares and the price at which to trade. Table 1 summarizes this information.

Table 1. Simulation details of each experiment.

Simulation Components	Experiment 1	Experiment 2
Number of runs at each level of algorithmic trading presence	50	10
Number of trading days per run	30 days	30 days
Number of price points recorded in each run	800 (equally spaced)	800 (equally spaced)
Number of traders	2000	200
Types of traders	Technical and fundamental, human, and algorithmic trading (AT)	Pairs traders (AT) and fundamental (human)

3.1. Initial Settings

We generate a stock price *history* by simulating a random initial price between \$5 and \$400. The series of daily prices is generated by drawing from a normal distribution whose mean is the previous day’s price. The final price in the 20-day price history is the *initial price* of the stock for the simulation. We also randomly generate the *fundamental value* of the stock, or how much the stock should actually be worth given the value of the company. The fundamental value of the stock on the first day of our simulation is a price uniformly chosen at random within $\pm 8\%$ of the initial price in the market. Then each day after that, the fundamental value is a random value within $\pm 8\%$ of the value the day before.

3.2. Traders’ Parameters

We model three trader strategies: fundamental analysis, technical analysis, and statistical arbitrage/pairs trading. All fundamental traders have the same strategy whether they are algorithmic traders or humans. The same is true for technical traders. Pairs traders are only simulated as algorithmic traders.

The difference between human traders and algorithmic traders is that algorithmic traders trade with higher frequency and lower latency than human traders. Frequency refers to how long it takes the trader to make a decision after the last decision. Latency refers to the time after a decision has been made and before the trade is executed. Therefore, the algorithmic traders can both make trade decisions more often and can execute their trades faster. In addition, algorithmic traders lack human judgement. An example is algorithmic pairs traders, who use no judgment regarding the true association between stocks and trade based solely on short-term price correlations.

Trader parameters are as follows. Human traders make their first decision in the fifth second of the simulation. They can execute their first trade at 10 s. Human trading frequency is between 95 and 105 s, i.e., after their last decision they wait a random value between 95 and 105 s before they make the next decision. Human latency is between 35 and 45 s. Algorithmic traders are much faster. They make their first decision within the first second and can execute in the next second. Then, latency and frequency are both one second so they can make a decision every second and always trade the second after that.

Finally, a risk parameter measures risk tolerance of traders; each trader starts with a budget that gets updated as traders execute trades and the random parameter measures the conditions prevailing

in the market at the time of trade such as investor sentiment, etc. All parameters are summarized in Table 2.

Table 2. Simulation Parameters.

Parameter and Basis for Setting the Value	Value
20-day history generated per day by drawing from a normal distribution with mean as the previous day price and with daily standard deviation of the S&P in 2006. Stock price history initial point, randomly generated:	\$5–\$400
Fundamental value of the stock	Within 8% of the price initially, thereafter it is within 8% of the fundamental value the day before
One trading day	12 h
Frequency, humans are slower to make a trade decision.	Start decision at 5th second after start of simulation for human trader; start decision at 0 s after simulation starts for AT. Thereafter humans decide to trade between every 95 to 105 s and AT every second.
Latency, humans take longer to execute trade once they make the decision to trade.	1st trade is executed at 10 s for humans and at 2 s for AT. Thereafter every trade is executed after 35–45 s of trade decision for humans and every second after trade decision for AT.
Risk, higher number indicates the capacity to take greater risk.	Set randomly for each buyer between 1 and 10 and for each seller between 0.5 and 1.5.
Initial budget for each trader	\$360
Percentage change in stock price that triggers a trade in the paired stock	One percent or more in the past 20 s
Perceived value	<ul style="list-style-type: none"> • 20-day moving average for technical traders • Within 30% of fundamental value for fundamental traders, • Max price (if price is falling), or min price (if price is rising) over past 20 s for pairs traders
Base volume bought	(Budget/Current Price)/5
Random	Set for each buy trade between 1 and 2 and set for each sell trade between 0.85 and 1.15

3.3. The Strategies

To approximate the behavior of traders in real markets, we implemented trading strategies based on commonly understood behaviors of traders in the market.

Fundamental traders trade based on what they believe to be the fundamental value of a stock. In our simulation at any given time, each fundamental trader is endowed with a random perceived fundamental value that is within 30% of the actual fundamental value. Differences between the perceived value and current price, drive trade decisions. If current price is greater than value, the trader sells the stock, and if current price is lower, she buys.

Our simulations are based on practically used fundamental trading strategies from the market. A few examples will help illustrate that fundamental traders’ decision-making process is in fact very similar to the simulations in our paper. The basic strategy in the paper consists of two steps: (1) arriving at a perceived value for the asset; and (2) trading when the asset deviates from this perceived value. This is paralleled in the real world by common strategies used by fundamental traders, for example, the commonly employed fundamental strategy of buying shares when they fall below a certain percentage of the perceived value, or when the stock is sold once it reaches its perceived value (Regan 2019). Another example is when fundamental traders employ “trailing stops”, where they sell any stock that falls by a certain percentage.

Note that fundamentals changing news may arrive only sporadically. However, actual fundamental trading frequency depends on how often, and in which direction, the price deviates from the trader determined perceived value. Price deviations away from perceived value are determined by the actual

trades in the market, by all types of traders. Therefore, a fundamental trader actually trades based on the number of times the price changes in a significant way with respect to the perceived value, and not necessarily based on the frequency of fundamental news arrival. For example, some news may not even result in any trade, if the change of the perceived value is such that there is no incentive for the trader to change their position in the market. On the other hand, even without any news arrival, movement in market prices, due to other traders' activities, may lead the fundamental trader to trade.

Technical traders' strategy is to consider the difference between their perceived value and the current price of the stock. They use the 20-day moving average (MA) price of the stock, which updates at every second based on the last 20 days' worth of seconds, as their perceived value for the stock. Trading decision is based on the difference between this perceived value and the current price. Specifically, a rise in current price above the MA triggers a buy order and the opposite triggers a sell order.

Following a statistical arbitrage strategy, a pairs trader trades one stock based on the movements of the correlated stock. Our goal is to determine the impact on the market when a pairs trader mistakenly believes two stocks are positively correlated, and begins trading them as though they are paired, even though they are fundamentally unrelated. So, our pairs trader will assume two unrelated stocks are positively and closely correlated. Then, if one stock moves a significant amount ($\pm 1\%$) since the trader checked last (20 s ago) the trader will buy if there was a rise and sell if there was a fall, the other stock at an appropriate price and volume.

Our simulations therefore mimic one of the most common algorithmic trading strategies: statistical arbitrage (Brogaard 2010; Zhang 2010; Froot et al. 1992). Statistical arbitrage strategies use short-term correlations among security prices to make short-term price predictions and trade to profit from these predictions.

3.4. Trading

Once each trader in our simulation makes the decision to trade, they all share a similar framework for deciding volume and price of trade. We use trade price to refer to the dollar value at which a trader wishes to trade the stock, and trade volume to refer to the number of shares that the trader wants to trade. The decision to trade begins with the consideration of each trader's estimate of the movement of the current price away from the perceived value. We define movement as the absolute value of the following ratio:

$$\text{movement} = \left| \frac{\text{perceived value} - \text{current price}}{\text{current price}} \right|, \quad (1)$$

where the perceived value for the fundamental trader is the perceived fundamental value, for the technical trader is the 20-day MA, and for the statistical arbitrage/pairs trader is a maximum (if price is falling) or minimum price (if price is rising) of the related stock within the past 20 s. Thus, movement measures how far the stock is mispriced from the point of view of the individual trader (where the mispriced stock is the paired stock for the Pairs trader).

For a buy order, the volume is determined as follows:

$$\text{buy volume} = \text{base volume} \times \text{movement} \times \text{random} \times \text{risk}, \quad (2)$$

and the price at which a buy order is executed is as follows:

$$\text{buy price} = \text{current price} + \text{movement} \times \text{random} \times \text{risk}, \quad (3)$$

where base volume is an initial trade volume, and the parameter random accounts for unpredictable elements that may change from one trade to the next, like investor mood. Buy volume is subject to the trader's budget. If their budget is less than the cost of the volume of shares they wish to buy, then the volume is automatically truncated as needed. Random and risk considerations will determine the exact trade price in any given transaction.

For a sell order, the volume is determined as follows:

$$\text{sell volume} = \text{volume held} \times \text{movement} \times \text{random} \times \text{risk}, \quad (4)$$

and the price at which a buy order is executed is as follows:

$$\text{sell price} = \text{current price} - \text{movement} \times \text{random} \times \text{risk}, \quad (5)$$

where volume held is the number of shares of the stock that the trader currently owns. If the sell volume is greater than the amount currently held, then all the held shares are sold. Again, in this case as well, random and risk considerations will determine the exact trade price in any given transaction.

3.5. Evolution of the Price

We implement a standard order book. Our order book is made up of two ordered lists, one for buy orders, and one for sell orders. The sell order list is ordered from low to high and the buy ordered list is ordered from high to low. When an order is placed it gets put into one of these two lists. All experiments start with the randomly generated start price as outlined in Section 3.1. Once a trade occurs, the price in the market is updated to the latest price at which the trade occurred.

4. Experiments and Results

We use our simulated asset market to conduct two separate experiments to evaluate the impact of AT on (1) market liquidity; and (2) in the presence of statistical arbitrage/pairs trading, the relations between the price and the fundamental value of the stock. In each experiment, our strategy is to gradually increase the market presence of AT starting from 0% of the market where there are no AT to 100%, where all trade is conducted exclusively by AT. We evaluate the market quality of interest at each level of market presence of AT, thus studying how the market quality might change with the advent and progression of AT.

For each experiment, the results reported represent an average of a number of runs of the experiment, where each run comprises a fixed numbers of trading days. The market outcome of interest is recorded a number of times during each run, this outcome is then averaged for each run and presented in the results. The number of traders in each experiment remains constant while the mix of the types of traders is varied to evaluate changes to the market outcome.

The number of each type of trade, in our simulations, is proportional to the trading latency and the proportion of traders in each run. These differ in each set as algorithmic trading is increased as a proportion of the market (Table 1). The number of fundamental human trades is always smaller than algorithmic trades of all varieties due to two factors. First, this is because of longer latency of human fundamental traders. Second, as the number of fundamental human traders is progressively reduced in each experiment, the numbers of their trades fall. Third, the number of trades vary, due to the randomness of each run in terms of the initial price and the evolution of the price through subsequent trade.

4.1. Experiment 1: Liquidity

We use effective half-spread as our measure of liquidity (e.g., [Hendershott and Riordan 2011](#)). As half-spread decreases, liquidity increases. Half-spread is calculated as:

$$\text{Half spread} = (\text{sign}) \frac{(\text{trade price} - \text{midpoint})}{\text{midpoint}}, \quad (6)$$

where trade price is the price at which trade occurred, midpoint is the price halfway between the transacting bid and ask price, and sign is determined by how the trade was initiated. If a seller initiated the trade then sign = 1, if a buyer initiated the trade then sign = -1. We compute the half spread at

800 evenly distributed points in time over the course of every single 30-day run (see Table 1). At the end of each run, we compute the average half-spread over the recorded 800 values. Each point on the curve in Figure 1 is thus the average of 50 average half spreads (Table 1). We present the results with 95% confidence intervals. For the first point on the graph, our simulation has 1000 human technical traders and 1000 human fundamental traders, with no algorithmic traders. In each successive point of the graph, the proportion of algorithmic traders increases. The number of technical and fundamental traders is equal at all points.

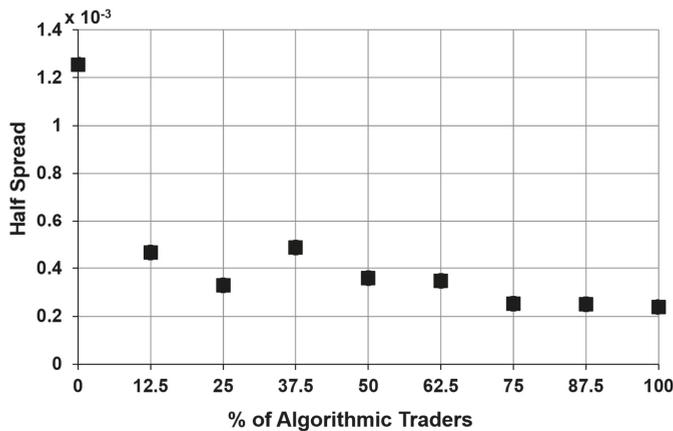


Figure 1. Impact of AT on market liquidity measured by the half-spread.

Results indicate that AT has a dramatic liquidity increasing effect early in its adoption. By the time AT reaches 10%, most of the benefit has already been gained. Further increase in the share of AT contributes very little to liquidity. This result indicates an interesting new hypothesis for the empirical literature: as AT presence increases in developing countries, where it is still only a fraction of the total market, empirical studies could estimate if there is a threshold where most of the benefit of AT already accrues to the country. This would provide practical insight into the timeline of AT impact on market quality in developing countries where AT adoption is still at a nascent stage.

4.2. Experiment 2: Statistical Arbitrage/Pairs Trading

Our goal is to determine the impact of statistical arbitrage/pairs trading. Statistical arbitrage is a strategy to profit by trading based on short-term correlations, while assuming that these correlations will persist. We simulate two fundamentally unrelated (low positive correlation) stocks. We introduce fundamental traders and pairs traders and measure the correlation of the prices and the correlation of the fundamental values of the stocks as the proportion of pairs traders increases.

The results in the form of the correlations between the two paired stocks are shown below in Figure 2. All details of the simulation are given in Table 1. The set-up is similar to Experiment 1 except for the types of traders. Figure 2 shows that the correlation of prices is significantly higher than the correlation in fundamental values of the two stocks. This is in line with empirical findings and anecdotal evidence that indicate AT leading to co-movements in market prices of assets (e.g., [Zweig 2010](#)), and price spillovers across assets (e.g., [Kelejian and Mukerji 2016](#)), which are above and beyond what fundamentals alone can explain.

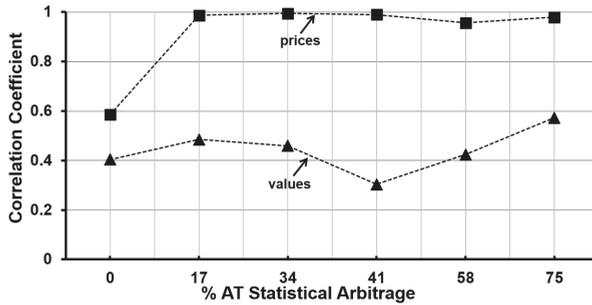


Figure 2. Impact of AT on relation between price and fundamentals.

To take a closer look at what unfolds over the course of a typical run, we present Figure 3, which shows what the market looks like when no Pairs traders are present.

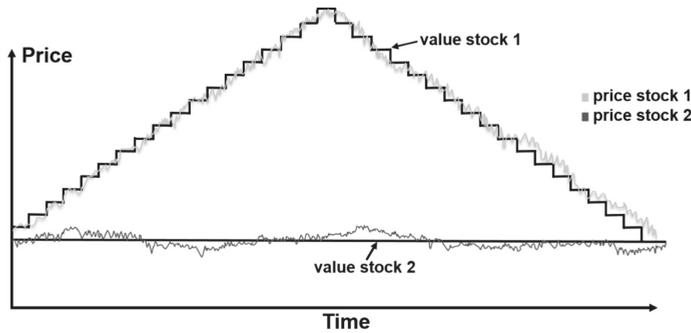


Figure 3. Single run of the simulation with no algorithmic traders in the market.

In this run, we control the fundamental values rather than randomly generating them. The first stock’s fundamental value is a step function that increases and then decreases. The second stock’s fundamental value line is the flat line. We see that when there are no Pairs traders, the prices stay near their corresponding fundamental values. In contrast, in Figure 4 we show the same simulation run now with the market comprising 50% algorithmic Pairs traders that trade on the assumption of positive price correlation. The prices in this case are far from their fundamental values and are highly correlated with one another.

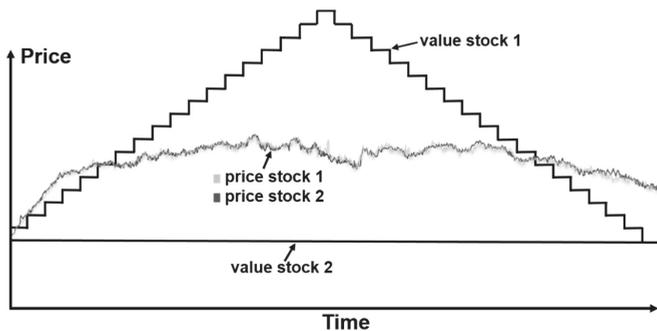


Figure 4. Single run of the simulation with 50% algorithmic pairs traders in the market.

This strong impact of AT on prices lends support to findings such as that the initial price impact of an algorithmic trade is larger than that of human orders (Hendershott and Riordan 2011; Hendershott et al. 2011) or that AT contribute significantly to the price discovery process (Brogaard 2010). Our contribution is to show the possibility that the price discovery process might be led astray, due to the strategy combined with the speed of the AT, as we see in Experiment 2.

5. Conclusions

In this paper, we have created a simulated asset market and modeled common AT strategies. We conducted two experiments in this market. Our contribution is the creation of a platform capable of studying the impact of AT trading on market quality directly. This expands our capability of evaluating AT beyond conventional empirical studies. It provides a method of cross-checking empirical findings, gaining deeper insights into them, and also potentially forming new hypotheses for further empirical research.

In Experiment 1, we investigated the impact of AT on market liquidity. We simulated a gradual rise in the participation of AT in a market with fundamental and technical traders of both AT and human types. We found that liquidity rose sharply as AT was introduced with most of the liquidity rise realized at just 10% AT participation. This leads to an interesting hypothesis for future empirical work to see if there is such an early threshold for reaping the liquidity benefits of AT. This could hold important implications for developing countries with nascent AT participation.

In Experiment 2 we investigated the impact of AT on price discovery. We simulated a rise in algorithmic pairs trading that apply statistical arbitrage to make profits based on short-term price correlations. In this experiment, we found high price correlations between the pair traded, even when fundamental values had very little correlation. This supports some findings in the empirical literature that indicate AT leading to price movements away from fundamentals.

Author Contributions: Conceptualization, P.M.; Methodology, P.M. and C.C.; Software, C.C., T.W., and B.X.; Validation, P.M., C.C., T.W. and B.X.; Formal Analysis, P.M., C.C., T.W. and B.X.; Investigation, P.M., C.C., T.W. and B.X.; Resources, P.M., and C.C.; Data Curation, P.M., C.C., T.W. and B.X.; Writing-Original Draft Preparation, P.M., C.C., T.W. and B.X.; Writing-Review & Editing, P.M., C.C. and T.W.; Visualization, P.M., C.C., T.W. and B.X.; Supervision, P.M., and C.C.; Project Administration, P.M. and C.C.; Funding Acquisition, P.M. and C.C.

Funding: This research was funded by research grants at Connecticut College.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Arthur, W. Brian, John H. Holland, Blake LeBaron, Richard G. Palmer, and Paul Tayler. 1996. Asset Pricing under Endogenous Expectation in an Artificial Stock Market. Working Papers No. 96-12-093, Social Systems, Madison, WI, USA.
- Biais, Bruno, and Paul Woolley. 2011. High Frequency Trading. Working Paper, Toulouse School of Economics, Toulouse, France.
- Brogaard, Jonathan. 2010. High Frequency Trading and Its Impact on Market Quality. Working Paper, Northwestern University, Evanston, IL, USA.
- Brogaard, Jonathan, Terrence Hendershott, and Ryan Riordan. 2014. High-frequency trading and price discovery. *The Review of Financial Studies* 27: 2267–306. [[CrossRef](#)]
- Chaboud, Alain P., Benjamin Chiquoine, Erik Hjalmarsson, and Clara Vega. 2014. Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance* 69: 2045–84. [[CrossRef](#)]
- Foucault, Thierry, and Albert J. Menkveld. 2008. Competition for Order Flow and Smart Order Routing Systems. *The Journal of Finance* 63: 119–58. [[CrossRef](#)]
- Froot, Kenneth A., David S. Scharfstein, and Jeremy C. Stein. 1992. Herd on the street: Informational inefficiencies in a market with short-term speculation. *The Journal of Finance* 47: 1461–84. [[CrossRef](#)]
- Hasbrouck, Joel, and Gideon Saar. 2013. Low-Latency Trading. *Journal of Financial Markets* 16: 646–79. [[CrossRef](#)]

- Hendershott, Terrence, and Ryan Riordan. 2011. Algorithmic Trading and Information. Working Paper, University of California, Berkeley, CA, USA.
- Hendershott, Terrence, Charles M. Jones, and Albert J. Menkveld. 2011. Does Algorithmic Trading Improve Liquidity? *The Journal of Finance* 66: 1–34. [CrossRef]
- Kearns, Michael, Alex Kulesza, and Yuriy Nevmyvaka. 2010. Empirical Limitations on High Frequency Trading Profitability. Available online: <http://ssrn.com/abstract=1678758> (accessed on 15 August 2011).
- Kelejian, Harry H., and Purba Mukerji. 2016. Does high frequency algorithmic trading matter for non-AT investors? *Research in International Business and Finance* 37: 78–92. [CrossRef]
- McCrank, John. 2015. 'Flash Crash' Trader's Alleged Fraud a Common Market Occurrence. New York: Reuters.
- Menkveld, Albert J. 2013. High frequency trading and the new market makers. *Journal of Financial Markets* 16: 712–40. [CrossRef]
- Philips, Matthew. 2013. How the Robots Lost: High Frequency Trading's Rise and Fall. *Bloomberg Businessweek*, June 6.
- Regan, Michael P. 2019. Buying is Easy. Selling is Hard. *Bloomberg Businessweek*, February 18, 22–23.
- Riordan, Ryan, and Andreas Storkenmaier. 2011. Optical Illusions: The Effects of Exchange System Latency on Liquidity. Working Paper, University of Karlsruhe, Karlsruhe, Germany.
- Sherry, Alok Mohan. 2017. *A Study on Algorithm Trading/High Frequency Trading in the Indian Capital Market*. New Delhi: National Institute of Financial Management, Ministry of Finance, Government of India.
- Sohel Azad, A. S. M., Saad Azmat, Victor Fang, and Piyadasa Edirisuriya. 2014. Unchecked Manipulations, Price Volume Relationship and Market Efficiency: Evidence from Emerging Markets. *Review of International Business and Finance* 30: 51–71. [CrossRef]
- U.S. Commodities Futures Trading Commission, and U.S. Securities and Exchange Commission. 2010. *Findings Regarding the Market Events of May 6, 2010*. Washington: United States Securities and Exchange Commission.
- Zhang, Frank. 2010. High-Frequency Trading, Stock Volatility, and Price Discovery. Available online: <http://ssrn.com/abstract=1691679> (accessed on 26 September 2013).
- Zweig, Jason. 2010. The Intelligent Investor: Why Your Stock Portfolio Is Acting Like a Commodity Basket. *The Wall Street Journal*, November 20.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Article

Efficient Numerical Pricing of American Call Options Using Symmetry Arguments

Lars Stentoft ^{1,2}

¹ Department of Economics, University of Western Ontario, London, ON N6A 5C2, Canada; lars.stentoft@uwo.ca

² Department of Statistical and Actuarial Sciences, University of Western Ontario, London, ON N6A 5B7, Canada

Received: 12 March 2019; Accepted: 4 April 2019; Published: 9 April 2019

Abstract: This paper demonstrates that it is possible to improve significantly on the estimated call prices obtained with the regression and simulation-based least-squares Monte Carlo method by using put-call symmetry. The results show that, for a large sample of options with characteristics of relevance in real-life applications, the symmetric method performs much better on average than the regular pricing method, is the best method for most of the options, never performs poorly and, as a result, is extremely efficient compared to the optimal, but unfeasible method that picks the method with the smallest Root Mean Squared Error (RMSE). A simple classification method is proposed that, by optimally selecting among estimates from the symmetric method with a reasonably small order used in the polynomial approximation, achieves a relative efficiency of more than 98%. The relative importance of using the symmetric method increases with option maturity and with asset volatility. Using the symmetric method to price, for example, real options, many of which are call options with long maturities on volatile assets, for example energy, could therefore improve the estimates significantly by decreasing their bias and RMSE by orders of magnitude.

Keywords: least-squares Monte Carlo; put-call symmetry; regression; simulation

JEL Classification: C15; G12; G13

1. Introduction

In a paper published some 20 years ago, [McDonald and Schroder \(1998\)](#) demonstrated that when the price of the underlying asset is governed by a Geometric Brownian Motion (GBM), the price of a call option with underlying asset price S , strike price K , interest rate r and dividend yield d is equal to the price of an otherwise identical put option with asset price K , strike price S , interest rate d and dividend yield r . The result for the GBM case has since been generalized to more realistic dynamics in [\(Schroder 1999\)](#), among others, and essentially, some version of this Put-Call Symmetry (PCS), potentially with other fundamental parameters changed accordingly, will hold for virtually all the models that have been considered in the existing literature on option pricing, for options with several different payoffs and which are written on multiple assets.¹

In this paper, we show that this simple result can be used to improve on one of today's state-of-the-art numerical option pricing methods, the well-known Least-Squares Monte-Carlo (LSMC)

¹ For example, PCS also holds in the stochastic volatility model of [Heston \(1993\)](#) when the parameters of the volatility process and the correlation are changed appropriately. See, e.g., [Battauz et al. \(2014\)](#) for the exact specification, [Grabbe \(1983\)](#) for an intuitive explanation of how to derive the relationship using options on foreign exchange and [Detemple \(2001\)](#) for extensions to derivatives on multiple assets.

method proposed by Longstaff and Schwartz (2001). In particular, we show that using PCS with LSMC results in estimates that are much less biased and have significantly lower Root Mean Squared Error (RMSE) when pricing American call options for the set of options used in (Longstaff and Schwartz 2001) and for a very large sample of options with realistic characteristics. Using standard choices for the LSMC method, which we implement with $N = 100,000$ paths and a polynomial of order $L = 3$, we price options with different strike prices and maturities in a world with different values for the interest rate, dividend yield and volatility. For a large sample of 3125 different options, the average RMSE of the estimates obtained with the symmetric method is only 17% of the RMSE of the estimates from the regular method, and the symmetric estimates have smaller RMSEs for 88% of the options in the sample.

Our results show that the relative performance of the symmetric method, i.e., when call options are priced as put options using PCS, improves as the time to maturity and volatility increase. Moreover, using the symmetric method is most effective for options that are out of the money. The simple intuition for this result is that when option maturity is long and volatility is high, asset values along simulated paths may become “very” large and be spread out over a large interval. Large and widely-spread out asset values lead to poorly-conditioned cross-sectional regressions and this in turn results in poor approximations of the optimal early exercise strategy and precisely determining this strategy is most important for out of the money options. Widely-dispersed asset values also lead to estimates that have higher variance because of the spread out payoffs being discounted back to estimate the price.

The magnitude of the relative improvement obtained with the symmetric method depends on the choice of parameters used in the LSMC algorithm, that is the number of simulated paths, N , and the number of regressors used in the cross-sectional regression, L , in a non-trivial way. In particular, while it is well known (see for example Stentoft (2004b)) that the option price estimated with the LSMC converges to the true value when the number of paths and the number of regressors tend to infinity, this is of little use with finite choices of the number of paths, N , and the order of the polynomial used in the regression, L . However, even with the “worst possible” configuration for the symmetric method, which occurs when $N = 100,000$ and $L = 5$ where the symmetric method only performs the best for roughly 39% of the individual options, the average RMSE is much smaller than for the regular method and only 18% larger than what could have been obtained with an infeasible method that picks from the regular and symmetric method the one with the smallest RMSE.

One reason that the choice of polynomial is important is that the LSMC method mixes two types of biases: a low bias due to having to approximate the optimal stopping time with a finite degree polynomial and a high bias coming from using the same paths to determine the optimal early exercise strategy and to price the option, potentially leading to over fitting to the simulated paths. For example, the bias just happens to be somewhat smaller without symmetry, a value of -0.006 , than when using symmetry, a value of 0.010 when using $L = 5$ regressors with $N = 100,000$ paths. For all other choices of the number of paths with this number of regressors and when using other numbers of regressors with this number of simulated paths, the symmetric estimates are less biased. An easy way to control the bias is to conduct so-called out-of-sample pricing in which a new set of simulated paths is used to price the option instead of using the same set of paths used for determining the optimal early exercise strategy.

When using out-of-sample pricing, the relative importance of the symmetric method is even more striking. In particular, the symmetric method almost always, and in some cases for more than 99% of the individual options, has the lowest RMSE, and the average RMSE for the large sample of options is around 20% or less of what is obtained with the regular method for most configurations. The efficiency of the symmetric method, when compared to the infeasible optimal method, is extraordinary and in most cases above 99% across various values of the number of paths, N , and number of regressors, L , whereas the regular method only achieves an efficiency of around 25%. Finally, while it is difficult to pick the best method in general, in the case of out of sample pricing, we propose a simple classification algorithm that, by optimally selecting among estimates from the symmetric method with

a reasonably small order used in the polynomial approximation, achieves a relative efficiency of more than 98% compared to the infeasible method that minimizes the RMSE across all estimates.

As noted by [Detemple \(2001\)](#), PCS is a useful property of many option pricing models since it reduces the computational burden when implementing these model. Indeed, a consequence of the property is that the same numerical algorithm can be used to price put and call options and to determine their associated optimal exercise policy. Another benefit is that it reduces the dimensionality of the pricing problem for some payoff functions. Examples include exchange options or quanto options. PCS also provides useful insights about the economic relationship between derivatives contracts. Puts and calls, forward prices and discount bonds and exchange options and standard options are simple examples of derivatives that are theoretically closely connected by symmetry relations. Compared to this literature, our objective is somewhat different. In particular, though PCS can be used to demonstrate theoretically the convergence of a particular numerical scheme for call option pricing using results for put options, our interest here is primarily of a numerical nature, and the objective is to show that PCS can be used to improve significantly on the estimated call option prices obtained with a particular numerical scheme.²

Our findings and proposed method for selecting optimally the configuration to use for option pricing should have broad implications. In particular, we show that improvements are found for a very large sample of options with reasonable characteristics, and since the symmetric method never performs very poorly and simple classification methods can be used to achieve very high relative efficiency, there are strong arguments for always using the symmetric method to price call options. Moreover, our results show that the relative importance of using the symmetric method increases with option maturity and asset volatility, and using symmetry to price long-term options in high volatility situations thus improves massively on the price estimates. The LSMC method is routinely used to price real options, most of which are call options with long maturities on volatile assets, for example energy. We conjecture that pricing such options using the symmetric method could improve significantly on the estimates by decreasing their bias and RMSE by orders of magnitude.

The rest of this paper is organized as follows: In Section 2, we provide motivating results for the small sample of simple vanilla options from [Longstaff and Schwartz \(2001\)](#). In Section 3, we briefly introduce the use of simulation methods for American option pricing in general and discuss the implementation of the method proposed by [Longstaff and Schwartz \(2001\)](#) when combined with PCS. In Section 4, we perform a large-scale study on 3125 options, showing that our proposed method works extremely well. In Section 5, we conduct several robustness checks, and in Section 6, we propose a new metric for efficiency and suggest a method for choosing the optimal specification to use for option pricing. Section 7 offers concluding remarks.

2. Motivation

In this section, we present results for a set of options similar to those used in [Longstaff and Schwartz \(2001\)](#), but to illustrate the effect of PCS, we consider pricing call options instead of put options. In all cases, we use a current value of the stock of 40 and an interest rate and dividend rate of 6%. A non-zero dividend is needed to make the American call option pricing non-trivial and to have positive early exercise premia. Options range between being 10% In The Money (ITM) and Out Of The Money (OTM), have maturities of $T = 1$ or $T = 2$ years and have $J = 50$ early exercise possibilities per year. We also consider two levels of the volatility and set $\sigma = 20\%$ or $\sigma = 40\%$. The reported estimates are based on $I = 100$ independent simulations, each of

² The main parts of the paper present results for the simple Black–Scholes–Merton setup. The reason for this is obvious: we want to have fast and precise benchmark results available. Without these, it makes no sense to talk about one method being more efficient than another. Section 5, though, shows that these conclusions extend to other asset dynamics, like the stochastic volatility model of [Heston \(1993\)](#), and to options with other payoff functions, like options written on multivariate underlying assets.

which uses $N = 100,000$ paths and the first $L = 3$ weighted Laguerre polynomials and a constant term as regressors in the cross-sectional regressions. We assess model performance using $Bias = I^{-1} \sum_{i=1}^I (\hat{P}_i - P)$, $StDev = \sqrt{I^{-1} \sum_{i=1}^I (\hat{P}_i - \bar{P}_i)^2}$ and $RMSE = \sqrt{I^{-1} \sum_{i=1}^I (\hat{P}_i - P)^2}$, respectively, where P is the true option price, \hat{P}_i is the i th simulated price and \bar{P}_i the average model price.

2.1. Regular Call Option Prices

Table 1 shows the pricing results for the sample of call options. The first thing to notice is that the majority of the estimated prices, shown in Column 5, are close to the benchmark values provided by the binomial model, shown in Column 4, and the bias, shown in Column 6, is in most cases less than one cent. However, for the longer term options with high volatility, shown in the last five rows, biases are large and significant at all reasonable levels. The size of the bias increases with moneyness, and the ITM option has a bias of 15 cents. Moreover, for these options, the estimated price also has a very large standard deviation, *StDev*, shown in Column 7, and as a result, the *RMSE*, shown in Column 8, is very large. For example, the *RMSE* of the option with $K = 36$, $T = 2$ and $\sigma = 0.40$ is almost 40-times larger than the *RMSE* of any of the other deep ITM options.

The results in Table 1 may hint at why often only put options are studied: it is potentially difficult to price long maturity call options in high volatility settings using the LSMC method. However, in many situations where the LSMC method is used, e.g., for real option pricing, the options considered are exactly long maturity call options. Therefore, what could (and does) go wrong? The fact that the standard deviation of these estimates is larger by (almost) an order of magnitude than that of any of the shorter term options indicates that this is likely caused by numerical issues. This conjecture is further supported by the fact that the skewness and kurtosis of the independent simulations are very far away from what we would expect, i.e., zero skewness and no excess kurtosis, when using independent simulations.³

Therefore, why then would you have numerical issues? The LSMC method estimates the early exercise strategy by performing a series of cross-sectional regressions of future path-wise payoffs on transformations of the current values of the stock price for the paths that are in the money, and the most obvious explanation for the numerical issues arising is that these regressions “break down” in one way or another. In particular, the properties of the input to the regression are very different when pricing calls, where regressors are unbounded, compared to when pricing puts, where regressors are bounded above by the strike price. Thus, one may end up performing regressions with regressors that have very large numerical values, and the probability of this happening increases with maturity and volatility. Note that this issue does not vanish when increasing the number of simulated paths, N .

2.2. Call Options Priced by Symmetry

When pricing call options using the “symmetric” method, the regressions carried out to price the, now, put option may be expected to be better behaved. In particular, the independent variable and the regressors are now bounded above by the strike price when using only the paths that are in the money. Columns 10–13 of Table 1 show the resulting price estimates, which may be compared directly to the estimates from the “regular” method in Columns 5–8. The first thing to notice is that with this approach, the estimated prices for the long-term high volatility options are now much closer to the benchmark values, and in fact, none of them are statistically different from the benchmark values provided by the binomial model. Note that some of the biases, five to be precise, are slightly larger for the symmetric method than for the regular method, yet in all cases, they are very small.

³ Although the path-wise payoffs obtained with the LSMC method for a given Monte Carlo simulation are dependent and could be very far from normally distributed, the price estimates we report in the table are averages of $I = 100$ independent simulations and should therefore be normally distributed by a central limit theorem. The actual values for the skewness and excess kurtosis are not shown in the table, but are available upon request.

Table 1. Call option prices.

K	T	σ	BM			Regular Call			Rel.			Symmetric Call		
			Price	Bias	StDev	Price	Bias	StDev	RMSE	RMSE	Price	Bias	StDev	RMSE
36	1	0.20	5.2247	5.2236	-0.0011	(0.0074)	0.0013	1.44	5.2240	-0.0007	(0.0059)	0.0009		
38	1	0.20	4.0292	4.0278	-0.0014	(0.0088)	0.0016	1.09	4.0278	-0.0014	(0.0050)	0.0015		
40	1	0.20	3.0420	3.0417	-0.0003	(0.0106)	0.0011	1.21	3.0414	-0.0006	(0.0071)	0.0009		
42	1	0.20	2.2502	2.2508	0.0006	(0.0099)	0.0012	1.52	2.2503	0.0001	(0.0075)	0.0008		
44	1	0.20	1.6324	1.6331	0.0007	(0.0100)	0.0012	1.52	1.6328	0.0004	(0.0070)	0.0008		
36	1	0.40	7.8808	7.8777	-0.0031	(0.0225)	0.0038	1.92	7.8790	-0.0018	(0.0083)	0.0020		
38	1	0.40	6.9153	6.9147	-0.0006	(0.0244)	0.0025	1.30	6.9135	-0.0018	(0.0076)	0.0019		
40	1	0.40	6.0543	6.0542	-0.0001	(0.0259)	0.0026	2.24	6.0537	-0.0006	(0.0101)	0.0012		
42	1	0.40	5.2900	5.2905	0.0005	(0.0241)	0.0025	1.92	5.2901	0.0001	(0.0128)	0.0013		
44	1	0.40	4.6141	4.6161	0.0021	(0.0235)	0.0031	2.34	4.6145	0.0004	(0.0127)	0.0013		
36	2	0.20	6.0796	6.0771	-0.0025	(0.0103)	0.0027	2.40	6.0787	-0.0009	(0.0070)	0.0011		
38	2	0.20	5.0249	5.0232	0.0016	(0.0119)	0.0020	2.37	5.0244	0.0005	(0.0071)	0.0009		
40	2	0.20	4.1221	4.1210	0.0010	(0.0135)	0.0017	2.25	4.1218	0.0002	(0.0072)	0.0008		
42	2	0.20	3.3578	3.3577	0.0001	(0.0132)	0.0013	1.30	3.3572	0.0006	(0.0079)	0.0010		
44	2	0.20	2.7174	2.7182	0.0008	(0.0126)	0.0015	1.67	2.7173	0.0001	(0.0090)	0.0009		
36	2	0.40	9.7706	9.6202	0.1504	(0.1055)	0.1508	113.89	9.7700	0.0006	(0.0120)	0.0013		
38	2	0.40	8.9315	8.8415	0.0900	(0.0844)	0.0904	64.28	8.9306	0.0009	(0.0112)	0.0014		
40	2	0.40	8.1661	8.1128	0.0533	(0.0634)	0.0537	38.56	8.1652	0.0009	(0.0106)	0.0014		
42	2	0.40	7.4684	7.4352	0.0332	(0.0496)	0.0336	30.01	7.4682	0.0002	(0.0111)	0.0011		
44	2	0.40	6.8323	6.8132	0.0191	(0.0411)	0.0195	15.19	6.8317	0.0006	(0.0116)	0.0013		

This table shows option prices for a set of benchmark call options from Longstaff and Schwartz (2001) using both the regular method and the symmetric method where the call option is priced as a put option. The values are based on $J = 100$ independent simulations, each of which uses $M = 100,000$ paths and the first $L = 3$ Laguerre polynomials and a constant term as regressors in the cross-sectional regressions using only the In The Money (ITM) paths. Benchmark values are from the Cox et al. (1979) binomial model with 25,000 steps and $J = 50$ early exercise possibilities per year.

However, not only is the bias of these estimates comparable across options, the standard deviation of the estimates is also similar across options. More importantly, the standard errors of the estimates are always lower than what is obtained with the regular method, and this is so even for the short-term options with low volatility in the first five rows. Across the 20 options, the regular method yields estimates with a standard error that is on average three-times larger, with the best case being roughly 26% worse (the option with $K = 36$, $T = 1$ and $\sigma = 0.20$), and the worst case having a standard deviation almost nine-times larger.

Because of the low bias and the much lower standard deviation, the RMSE of the call price estimates obtained using symmetry is much lower than that obtained when pricing the option with the regular method across the benchmark sample. For half of the options, the RMSE is less than half that obtained with the regular method when using the symmetric method. In the best case across the 20 options, the regular method is only 9% worse than the symmetry method; however, one would never do worse when pricing this set of call options using symmetry than with the regular method. This is a very strong argument for using symmetry to price call options.

3. Implementation

The first step in implementing any type of numerical algorithm to price American options is to assume that time can be discretized. Thus, we assume that the derivative considered may be exercised at J points in time. We specify the potential exercise points as $t_0 = 0 < t_1 \leq t_2 \leq \dots \leq t_J = T$, with t_0 and T corresponding to the current time and maturity of the option, respectively. An American option can be approximated by increasing the number of exercise points J , and a European option can be valued by setting $J = 1$. We assume a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ equipped with a discrete filtration $(\mathcal{F}(t_j))_{j=0}^J$. The derivative's value depends on one or more underlying assets, which are modelled using a Markovian process, with state variables $(X(t_j))_{j=0}^J$ adapted to the filtration and with $X(0) = x$ known. We denote by $(Z(t_j))_{j=0}^J$ an adapted payoff process for the derivative satisfying $Z(t_j) = \pi(X(t_j), t_j)$ for a suitable function $\pi(\cdot, \cdot)$, which is assumed to be square integrable. Following, e.g., Karatzas (1988) and Duffie (1996), in the absence of arbitrage, we can specify the American option price as:

$$P(X(0) = x) = \max_{\tau(t_1) \in \mathcal{T}(t_1)} E[Z(\tau) | X(0)], \tag{1}$$

where $\mathcal{T}(t_j)$ denotes the set of all stopping times with values in $\{t_j, \dots, t_J\}$ and where it is therefore implicitly assumed that the option cannot be exercised at time t_0 .

In the literature, the problem of calculating the American option price in Equation (1), i.e., with $J > 1$, is referred to as a discrete time optimal stopping time problem. The preferred way to solve such problems is to use the dynamic programming principle. Intuitively, this procedure can be motivated by considering the choice faced by the option holder at time t_j : either to exercise the option immediately or to continue to hold the option until the next period. Obviously, at any time, the optimal choice will be to exercise immediately if the value of this is positive and larger than the expected payoff from holding the option until the next period and behaving optimally from there on forward. To fix notation, in the following, we let $V(X(t_j))$ denote the value of the option for state variables X at a time t_j prior to expiration. We define $F(X(t_j)) \equiv E[Z(\tau(t_{j+1})) | X(t_j)]$ as the expected conditional payoff, where $\tau(t_{j+1})$ is the optimal stopping time. It follows that:

$$V(X(t_j)) = \max(Z(t_j), F(X(t_j))), \tag{2}$$

and it is easily seen that it is possible to derive the optimal stopping time iteratively using the following algorithm:

$$\begin{cases} \tau(t_j) = T \\ \tau(t_j) = t_j \mathbf{1}_{\{Z(t_j) \geq F(X(t_j))\}} + \tau(t_{k+1}) \mathbf{1}_{\{Z(t_j) < F(X(t_j))\}}, \quad 1 < j \leq J-1. \end{cases} \tag{3}$$

Based on this, the value of the option in Equation (1) can be calculated as:

$$P(X(0) = x) = E[Z(\tau(t_1)) | X(0)]. \tag{4}$$

The backward induction theorem of [Chow et al. \(1971\)](#) (Theorem 3.2) provides the theoretical foundation for the algorithm in Equation (3) and establishes the optimality of the derived stopping time and the resulting price estimate in Equation (4).

3.1. Simulation and Regression Methods

The idea behind using simulation for option pricing is quite simple and involves estimating expected values and therefore option prices by an average of a number of random draws. However, when the option is American, one needs to determine simultaneously the optimal early exercise strategy, and this complicates matters. In particular, it is generally not possible to implement the exact algorithm in Equation (3) because the conditional expectations are unknown, and therefore, the price estimate in Equation (4) is infeasible. Instead, an approximate algorithm is needed. Because conditional expectations can be represented as a countable linear combination of basis functions, we may write $F(X(t_j)) = \sum_{l=0}^{\infty} \phi_l(X(t_j)) c_l(t_j)$, where $\{\phi_l(\cdot)\}_{l=0}^{\infty}$ form a basis.⁴ In order to make this operational, we further assume that it is possible to approximate well the conditional expectation function by using the first $L + 1$ terms such that $F(X(t_j)) \approx F_L(X(t_j)) = \sum_{l=0}^L \phi_l(X(t_j)) c_l(t_j)$ and that we can obtain an estimate of this function by:

$$\hat{F}_L^N(X(t_j)) = \sum_{l=0}^L \phi_l(X(t_j)) \hat{c}_l^N(t_j), \tag{5}$$

where $\hat{c}_l^N(t_j)$ are approximated or estimated using $N \geq L$ simulated paths. Based on the estimate in Equation (5), we can derive an estimate of the optimal stopping time from:

$$\begin{cases} \hat{\tau}_L^N(t_j) = T \\ \hat{\tau}_L^N(t_j) = t_j \mathbf{1}_{\{Z(t_j) \geq \hat{F}_L^N(X(t_j))\}} + \hat{\tau}_L^N(t_{k+1}) \mathbf{1}_{\{Z(t_j) < \hat{F}_L^N(X(t_j))\}}, \quad 1 < j \leq J-1. \end{cases} \tag{6}$$

From the algorithm in Equation (6), a natural estimate of the option value in Equation (4) is given by:

$$\hat{P}_L^N(X(0) = x) = \frac{1}{N} \sum_{n=1}^N Z(n, \hat{\tau}_L^N(1, n)), \tag{7}$$

where $Z(n, \hat{\tau}_L^N(1, n))$ is the payoff from exercising the option at the optimal stopping time $\hat{\tau}_L^N(1, n)$ determined for path n according to Equation (6).

3.2. Implementation of the LSMC Method

When implementing the method outlined above, one has to choose at least two things: how to generate the data, the simulated state variables, and how to approximate the value function, that

⁴ This is justified when approximating elements of the L^2 space of square-integrable functions relative to some measure. Since L^2 is a Hilbert space, it has a countable orthonormal basis (see, e.g., [Royden 1988](#)).

is how to estimate the parameters in the approximation. The key contribution of Longstaff and Schwartz (2001) is to suggest that the coefficients in the approximation of the continuation value, $\hat{c}_t^N(t_j)$, can be estimated in a simple cross-sectional ordinary linear (OLS) regression, where the independent variable is the discounted path-wise future payoff and the dependent variables are functions of the current state variables. In this paper, we propose to merge the LSMC method with PCS and use the symmetric method when pricing call options. Thus, instead of simulating paths from a dynamic model with a risk-free rate of r and dividend yield of d , we simulate from the same dynamic model, but with a risk-free rate of d and dividend yield of r , and instead of pricing the option as a call option with a strike price of K and a current value of the underlying asset of S , we price the option as if it had been a put option with a strike price of S and a current value of the underlying asset of K .⁵ These changes are simple to make and involve no extra computational complexity or changes to the numerical procedure. In fact, for consistency, it is important to note that we use the exact same numerical procedures for simulating the paths and to implement the cross-sectional regression.⁶

There are two very intuitive reasons why using the symmetric method to price call options may work better than when pricing the call option using the regular method. First, as explained above, in simulation-based methods, the option price, an expectation under the risk neutral measure, is approximated by the average of a number of random realizations of future payoffs, obtained from simulated values of the appropriate state variables. This mean obviously behaves better, and the estimator will have a smaller variance when the possible realizations are bounded, as they are in the case of the payoff of a put option, than when they are unbounded, as they are in the case of the payoff of a call option. Our numerical results for the benchmark options in Section 2 indeed showed that the standard deviation of the estimates is always lower when using the symmetric method than when using the regular method.⁷

Second, it is easier to approximate the continuation value when this is a bounded function on a bounded interval than when this is an unbounded function on an unbounded interval. In particular, theoretically, it is straightforward to design a robust approximation scheme for the continuation value of a put option using only the simulated paths that are in the money.⁸ For call options, on the other hand, no general theoretical results exist to justify that this is in fact feasible, and though numerical schemes are available and polynomial families that have nice properties can be used, approximating the continuation value is likely much more complicated. A further complication with the continuation value of a call option is that this is bounded below by the exercise value for large values of the underlying asset and thus asymptotically linear in the stock value. It is obviously difficult to approximate a function with these characteristics.

4. Results

The motivating example in Section 2 clearly demonstrated that there might be significant value to pricing call options as put options using symmetry properties when using the LSMC algorithm. To test this further, we now price a large sample of call options with five different strike prices, $K = [90, 95, 100, 105, 110]$, maturities, $T = [0.5, 1, 2, 3, 5]$ years, interest rates, $r = [0.0\%, 2.5\%, 5.0\%, 7.5\%, 10\%]$, dividend yields, $d = [0.0\%, 2.5\%, 5.0\%, 7.5\%, 10\%]$, and volatilities, $\sigma = [10\%, 20\%, 30\%, 40\%, 50\%]$, for a total of $5 \times 5 \times 5 \times 5 \times 5 = 3,125$ options. This sample arguably

⁵ For now, we maintain the assumption that dynamics are governed by simple geometric Brownian motion. However, our results generalize to other models for which PCS holds, as we demonstrate in Section 5.

⁶ We deal with the difference in, for example, the payoff when exercising the option by using negative values for the strike price and the stock prices for put options since $Z_{Call}(S, K) = \max(S - K, 0) = \max((-K) - (-S), 0) = Z_{Put}(-S, -K)$.

⁷ Unreported results, available upon request, show that this generalizes to the much larger sample of options we consider in Section 4.

⁸ This follows from the Weierstrass approximation theorem, which states that every continuous function defined on a closed interval can be uniformly approximated as closely as desired by a polynomial function.

spans most of the important cases one would come across in real-life applications of option pricing. We first consider performance for various numbers of exercise possibilities J and across option characteristics, i.e., K and T . Next, we consider performance across model parameters, i.e., r , d and σ . Benchmark values are from the Cox et al. (1979) binomial model with 25,000 steps and J early exercise possibilities.

In this section, we use a slightly different setup for the LSMC algorithm in that we use monomials as regressors and we use simple “plain vanilla” Monte Carlo simulation. We choose monomials instead of Laguerre polynomials because they are simpler and faster to use. We choose a plain Monte Carlo simulation without any variance reduction techniques such that our results are not potentially dependent on a particular variance reduction method. We again report results with the LSMC method using $I = 100$ independent simulations with $N = 100,000$ paths and $L = 3$ regressors. We assess model performance using the $Bias = I^{-1} \sum_{i=1}^I (\hat{P}_i - P)$ and $RMSE = \sqrt{I^{-1} \sum_{i=1}^I (\hat{P}_i - P)^2}$ error metrics where \hat{P}_i is the i th simulated estimate of the price P . Since we cannot report all the individual errors, we report average errors instead. We also consider the fraction of options for which the regular and symmetric method have the highest bias and RMSE, respectively.⁹

4.1. Performance across Option Characteristics

Table 2 reports results for our benchmark implementation of the LSMC method for the large sample of call options considering different numbers of total early exercises, J , constant across the maturity, from $J = 10$ to $J = 200$ (a close approximation to the continuously-exercisable American option) and across different strike prices, K , and different maturities, T , for options with $J = 50$ early exercises. Figure 1 plots the relative performance of the symmetric method compared to the regular method for the four aggregate error metrics across these three dimensions.

Panel A of Table 2 first of all shows that using the regular method for this sample of options leads to significantly low biased price estimates. For example, when considering the case with $J = 50$ exercise times, the average bias is almost six cents with this method, whereas it is less than a cent if the symmetric method is used leading to an average improvement of $1 - |-0.0574 / -0.0022| = 96.22\%$. The improvement in performance with the symmetric approach is large also for the RMSE. Moreover, the improvement in performance is not only large on average, but also across most of the options, as the counting metrics show. In particular, improvements occur for 83.90% and 88.19% of the options in terms of bias and the RMSE, respectively. Figure 1a shows that the improved performance is not limited to a particular choice of the number of early exercise possibilities, J , and improvements are found for all the reported values of J . Across the number of early exercises, the figure does indicate that the relative performance improves rapidly with J when there are only a small number of exercise possibilities. Once J reaches 50 or 100, the effect in terms of RMSE tapers off, and the relative improvement in performance does not change much when increasing the number of early exercise points further.¹⁰

Panel B of Table 2 shows the results across moneyness and demonstrates that the absolute errors of both methods decrease when the strike price increases. In terms of the counting metrics the panel shows that the symmetric method has the smallest errors for 98.4% of the out of the money options. For in the money options, where determining the early exercise strategy is of less importance, the improvement is relatively smaller, though the symmetric method continues to yield more precise price estimates on average and for the majority of the options. In relative terms, Figure 1b shows that the symmetric method performs better than the regular method across all strike prices. The relative

⁹ Using the fraction of times a given method has the highest error metric ensures, as is the case with the bias and RMSE error metrics, that lower numbers are better.

¹⁰ Note, though, that, e.g., the bias of both the regular and symmetric method increases in absolute terms when increasing the number of exercise points. This is likely related to the fact that dependence is introduced between the paths in the LSMC method because of the cross-sectional regression, and this dependence “accumulates” as we go backwards in time in the algorithm and becomes more and more important as the number of early exercise possibilities increases.

performance is best for options with low strike prices, i.e., call options that are out of the money, and for these options, the relative performance in terms of the counting metrics is quite extraordinary.

Finally, Panel C of Table 2 shows the results across maturity and documents clear and significant improvements in the relative performance of the symmetric method for all metrics when maturity increases. It is noteworthy that the symmetric method actually leads to more precise, in terms of RMSE, estimates for all subcategories. In terms of the counting metrics, the symmetric method also largely outperforms the regular method and leads to prices estimated with smaller errors in at least 72.16% of the cases, the worst relative performance being for the shortest maturity options. In relative terms, Figure 1c shows that the symmetric method performs better than the regular method across all maturities and error metrics. In fact, for the majority of the categories, that is for options with maturity of $T = 2$ years or more, the symmetric method leads to lower RMSE for at least nine out of 10 of the options. Thus, the results from Section 2 hold true in general for a much larger sample of options.

Table 2. Pricing errors across option characteristics.

Panel A: Across Early Exercise Points J								
J	Absolute Metrics				Counting Metrics			
	Regular Call		Symmetric Call		Regular Call		Symmetric Call	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
10	-0.0066	0.0169	0.0022	0.0063	0.6525	0.7878	0.3475	0.2122
25	-0.0310	0.0351	0.0038	0.0086	0.7251	0.7894	0.2749	0.2106
50	-0.0574	0.0602	-0.0022	0.0104	0.8390	0.8819	0.1610	0.1181
100	-0.0639	0.0660	-0.0090	0.0138	0.9222	0.9683	0.0778	0.0317
200	-0.0922	0.0963	-0.0207	0.0222	0.8829	0.9014	0.1171	0.0986

Panel B: Across Strike Prices K								
K	Absolute Metrics				Counting Metrics			
	Regular Call		Symmetric Call		Regular Call		Symmetric Call	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
90	-0.0682	0.0707	-0.0055	0.0112	0.9680	0.9840	0.0320	0.0160
95	-0.0631	0.0657	-0.0037	0.0109	0.9680	0.9936	0.0320	0.0064
100	-0.0574	0.0602	-0.0019	0.0104	0.8560	0.9232	0.1440	0.0768
105	-0.0521	0.0550	-0.0005	0.0100	0.7520	0.8128	0.2480	0.1872
110	-0.0464	0.0495	0.0007	0.0096	0.6512	0.6960	0.3488	0.3040

Panel C: Across Maturity T								
T	Absolute Metrics				Counting Metrics			
	Regular Call		Symmetric Call		Regular Call		Symmetric Call	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.5	-0.0211	0.0220	-0.0050	0.0107	0.7216	0.7504	0.2784	0.2496
1	-0.0279	0.0293	-0.0032	0.0115	0.7936	0.8432	0.2064	0.1568
2	-0.0392	0.0414	-0.0011	0.0110	0.8560	0.9056	0.1440	0.0944
3	-0.0525	0.0553	-0.0004	0.0100	0.9008	0.9472	0.0992	0.0528
5	-0.1462	0.1531	-0.0011	0.0089	0.9232	0.9632	0.0768	0.0368

This table shows pricing errors for the regular and symmetric method for various numbers of exercise possibilities, J , strike prices, K , and maturities, T . Results are based on $I = 100$ independent simulations with $N = 100,000$ paths and $L = 3$ regressors. In each panel, we report results for the bias and RMSE in terms of the average metrics and counting metrics, i.e., the fraction of times a given method has the highest error metric. Panels B and C use results for $J = 50$ early exercise points only.

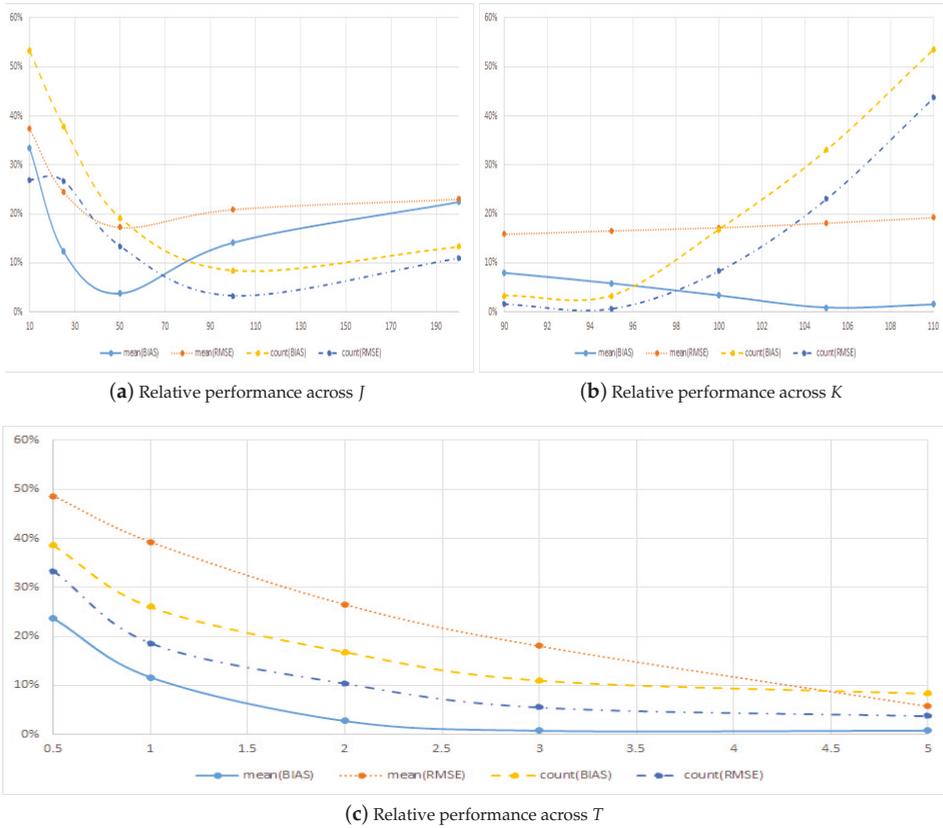


Figure 1. Relative pricing performance across option characteristics. This figure plots the relative performance of the symmetric method compared to the regular method across the number of early exercises, J , strike price, K , and maturity, T . Panels B and C use results for $J = 50$ early exercise points only.

4.2. Performance across Model Parameters

We now consider the relative performance of the two methods for some interesting subgroups of model parameters like the interest rate r , the dividend yield d and the volatility of the underlying asset σ . Table 3 shows the results across interest rate, r , dividend yield, d , and volatility, σ , for our large sample of options. Figure 2 plots the relative performance of the symmetric method compared to the regular method across these three dimensions.

Table 3. Pricing errors across model characteristics.

Panel A: Across Interest Rates r								
r	Absolute Metrics				Counting Metrics			
	Regular Call bias	Regular Call RMSE	Symmetric Call Bias	Symmetric Call RMSE	Regular Call Bias	Regular Call RMSE	Symmetric Call Bias	Symmetric Call RMSE
0%	-0.0560	0.0574	0.0025	0.0065	0.8752	0.8992	0.1248	0.1008
2.5%	-0.0656	0.0673	-0.0018	0.0108	0.8608	0.8784	0.1392	0.1216
5.0%	-0.0618	0.0643	-0.0038	0.0119	0.8304	0.8656	0.1696	0.1344
7.5%	-0.0553	0.0590	-0.0040	0.0118	0.8112	0.8688	0.1888	0.1312
10%	-0.0484	0.0530	-0.0038	0.0110	0.8176	0.8976	0.1824	0.1024

Table 3. Cont.

Panel B: Across Dividend Yields d								
d	Absolute Metrics				Counting Metrics			
	Regular Call		Symmetric Call		Regular Call		Symmetric Call	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0%	-0.0730	0.0789	-0.0093	0.0142	0.8368	0.9360	0.1632	0.0640
2.5%	-0.0555	0.0591	-0.0064	0.0147	0.8272	0.8864	0.1728	0.1136
5.0%	-0.0487	0.0509	-0.0009	0.0106	0.8432	0.8720	0.1568	0.1280
7.5%	-0.0527	0.0542	0.0030	0.0069	0.8496	0.8640	0.1504	0.1360
10%	-0.0571	0.0580	0.0027	0.0058	0.8384	0.8512	0.1616	0.1488

Panel C: Across Volatility Levels σ								
σ	Absolute Metrics				Counting Metrics			
	Regular Call		Symmetric Call		Regular Call		Symmetric Call	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
10%	-0.0034	0.0046	0.0010	0.0026	0.6464	0.7104	0.3536	0.2896
20%	-0.0162	0.0186	-0.0010	0.0075	0.7584	0.8512	0.2416	0.1488
30%	-0.0382	0.0411	-0.0025	0.0114	0.8848	0.9152	0.1152	0.0848
40%	-0.0694	0.0716	-0.0037	0.0143	0.9328	0.9552	0.0672	0.0448
50%	-0.1598	0.1652	-0.0047	0.0161	0.9728	0.9776	0.0272	0.0224

This table shows pricing errors for the regular and symmetric method for various numbers of interest rates r , dividend yields, d , and volatility level, σ . Results are based on $I = 100$ independent simulations with $N = 100,000$ paths and $L = 3$ regressors. In each panel, we report results for the bias and RMSE in terms of the average metrics and counting metrics, i.e., the fraction of times a given method has the highest error metric.

Figure 2a,b along with Panels A and B of Table 3 show that the relative improvement from using the symmetric method is large across all the possible values of interest rates and dividend yields. This holds both in terms of the average metrics and in terms of the number of options for which the symmetric method has the smallest error. In terms of absolute errors, the relative performance of the symmetric method decreases slightly when the interest rate increases, though the method produces estimates with errors that are very small and never above one third of the errors obtained with the regular method. When the dividend yield increases, the relative performance of the symmetric method increases somewhat. Note that the case with $d = 0$ is special since in this situation, the American call option should never be exercised early.

Figure 2c and Panel C of Table 3 document clear and significant improvements in the relative performance of the symmetric method for all metrics in absolute, as well as relative terms when volatility increases. It is noteworthy that the symmetric method actually leads to more precise, in terms of RMSE, estimates for all subcategories. In terms of the counting metrics, the symmetric method also largely outperforms the regular method and leads to price estimates with smaller errors in at least 64.64% of the cases, the worst relative performance being for the lowest volatility options. For the majority of the categories, that is for options with volatility of $\sigma = 0.30$ or more, the symmetric method leads to lower RMSE for at least nine out of 10 of the options. Thus, the results from Section 2 hold true in general for a much larger sample of options.

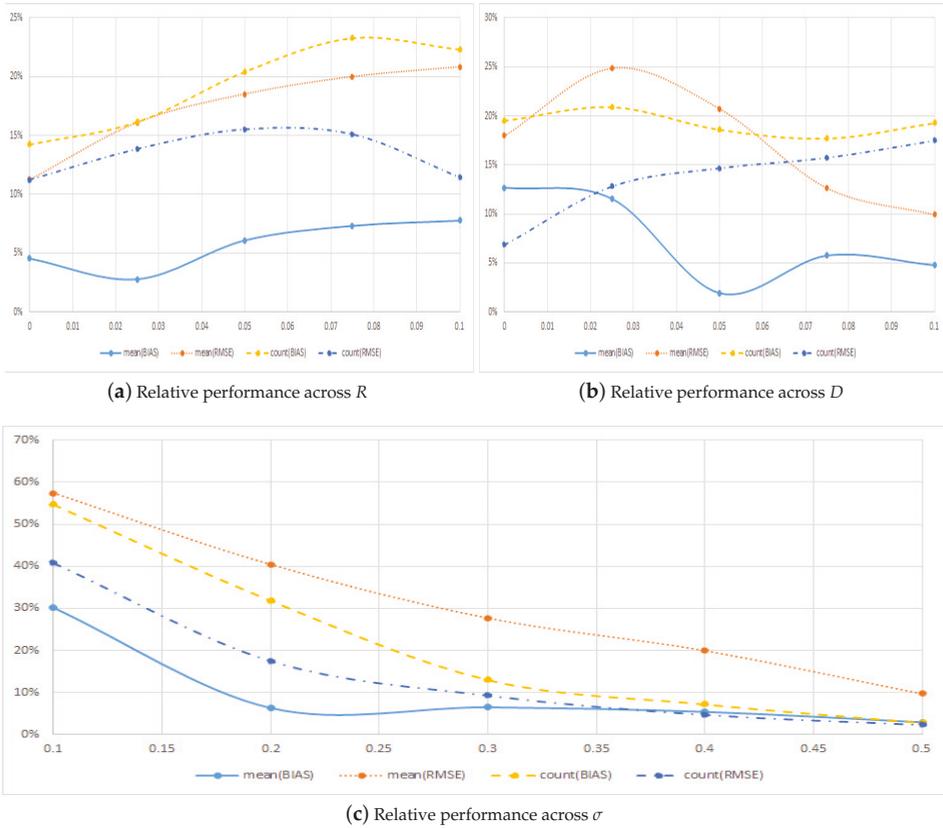


Figure 2. Relative pricing performance across model characteristics. This figure plots the relative performance of the symmetric method compared to the regular method across interest rates, r , dividend yields, d , and volatility levels, σ .

5. Robustness

The previous section provides strong evidence in favour of using symmetric pricing for call options. In this section, we examine the robustness of these results along two dimensions. We first examine the importance of the choice of the number of paths, N , and the number of regressors, L , used in the Monte Carlo simulation and whether or not our reported results are robust to using so-called out-of-sample pricing. Next, we examine the robustness of our results to using alternative option pricing models. Here, we consider the case with multiple underlying assets and the case in which the underlying asset follows the stochastic volatility model of [Heston \(1993\)](#).

5.1. Alternative Choices for the Number of Paths and Regressors

When implementing the LSMC method, one needs to choose the number of paths to simulate, N , and the number of regressors, L , to use in the cross-sectional regressions. While it is well known that the estimated prices converge to the true price when both N and L tend to infinity (see, e.g., [Stentoft \(2004b\)](#)), any real application involves choosing a finite number of paths and regressors. Table 4 shows the results across the number of simulated paths, N , and the number of regressors, L , for our large sample of options. Figure 3 plots the relative performance of the symmetric method compared to the regular method across these two dimensions.

Table 4. Pricing errors across algorithm characteristics.

Panel A: Across Number of Simulated Paths N								
N	Absolute Metrics				Counting Metrics			
	Regular Call		Symmetric Call		Regular Call		Symmetric Call	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
20,000	-0.0112	0.0426	0.0089	0.0178	0.5498	0.6915	0.4502	0.3085
50,000	-0.0428	0.0480	0.0038	0.0131	0.7130	0.8000	0.2870	0.2000
100,000	-0.0574	0.0602	-0.0022	0.0104	0.8390	0.8819	0.1610	0.1181
200,000	-0.0526	0.0540	-0.0080	0.0100	0.9322	0.9766	0.0678	0.0234
500,000	-0.0549	0.0556	-0.0120	0.0124	0.8749	0.8835	0.1251	0.1165

Panel B: Across Number of Regressors L								
L	Absolute Metrics				Counting Metrics			
	Regular Call		Symmetric Call		Regular Call		Symmetric Call	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
2	-0.0819	0.0832	-0.0330	0.0351	0.8608	0.8986	0.1392	0.1014
3	-0.0574	0.0602	-0.0022	0.0104	0.8390	0.8819	0.1610	0.1181
5	-0.0061	0.0201	0.0104	0.0117	0.2848	0.3901	0.7152	0.6099
9	0.0296	0.0328	0.0144	0.0154	0.6611	0.7123	0.3389	0.2877
15	0.0614	0.0625	0.0178	0.0187	0.7261	0.7555	0.2739	0.2445

This table shows pricing errors for the regular and symmetric method for various numbers of simulated paths N and number of regressors L . Results are based on $I = 100$ independent simulations. In each panel, we report results for the bias and RMSE in terms of the average metrics and counting metrics, i.e., the fraction of times a given method has the highest error metric.

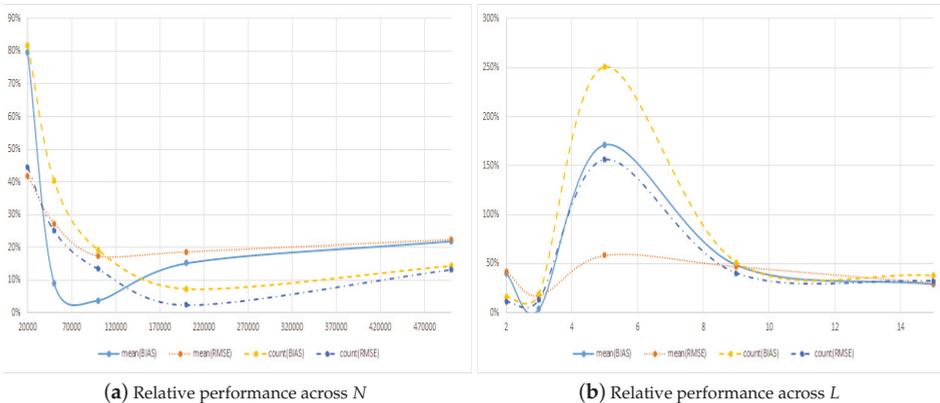


Figure 3. Relative pricing performance across algorithm characteristics. This figure plots the relative performance of the symmetric method compared to the regular method across the number of simulated paths, N , and number of regressors, L .

Panel A of Table 4 shows the results when increasing the number of simulated paths from $N = 20,000$ to $N = 500,000$, while keeping the number of regressors fixed at $L = 3$. In this case, we know that the methods converge to a low estimate of the true value, one that is based on using a rather rough approximation of the conditional expectation function used to determine the optimal early exercise. The table confirms this numerically in that the bias for both methods, regular as well as symmetric, becomes more negative with increasing N . Note also that the regular method always yields price estimates with a low bias on average even when using as low as $N = 20,000$ paths, whereas the symmetric method yields high biased estimates for low N .

When comparing the two methods, it is noteworthy, though, that in all cases, the absolute bias and the RMSE is lowest with the symmetric method, and this method consistently outperforms the regular method across all choices of N , as can be seen from Figure 3a. When it comes to the number of times the symmetric method has lower errors, a pattern very similar to what was seen when increasing T or σ is found.

Panel B of Table 4 shows the results when increasing the number of regressors from $L = 2$ to $L = 15$, while keeping the number of simulated paths fixed at $N = 100,000$. In this case, we know that, everything else equal, the estimated prices should increase as the approximation gets better and better, although this may eventually result in a high bias because of over fitting the function on a finite number of simulated paths. The table confirms this numerically in that the bias for both methods, regular as well as symmetric, becomes more positive with increasing L . The change is most dramatic for the regular method, which goes from having an average negative bias of close to eight cents to having an average positive bias of more than six cents.

When comparing the two methods, the table shows that the symmetric method almost always provides estimates with smaller errors than does the regular method across the choice of L . The exception to this is when using $L = 5$ and looking at the average bias, where the absolute value from using the regular method is half that of using the symmetric method. When it comes to the number of times the symmetric method has lower errors, a pattern very similar to what was observed previously is found. The main difference is that, for the first configuration of all the ones considered up to this point, a case occurs where the regular method on average provides estimates that are better in terms of the RMSE. Unsurprisingly, this happens when $L = 5$ for which 60.99% of the estimated regular prices have smaller errors.

When looking at Panel B of the table, it is noteworthy that the performance of the symmetric method is better for a small, i.e., $L \leq 3$, or a large, i.e., $L \geq 9$, choice of regressors. A similar, though less pronounced, non-linear relationship is found in Panel A of the table when the number of simulated paths is increased. Given this concave relationship, as a function of L , and convex relationship, as a function of N , in the relative performance, it is indeed possible that one could find a combination of L and N for which the regular method would outperform the symmetric method for our large sample of options. This though would be largely due to luck (or would require one to consider a large number of possible combinations) and as such is not of much help or relevance.

One reason that the choice of polynomial is important is that the LSMC method mixes two types of biases: a low bias due to having to approximate the optimal stopping time with a finite degree polynomial and a high bias stemming from using the same paths to determine the optimal early exercise strategy and to price the option, potentially leading to over fitting to the simulated paths. As a result of this, the bias just happens to be somewhat smaller without symmetry, a value of -0.006 , then when using symmetry, a value of 0.010 when using $L = 5$ regressors with $N = 100,000$ paths.¹¹ One easy way to control the sign of the bias is to conduct so-called out-of-sample pricing in which a new set of simulated paths is used to price the option instead of using the same set of paths that were used for determining the optimal early exercise strategy. Table 5 shows the results for the different configurations of N and L and, as expected and in line with the theory, shows that the bias of the estimates from the regular, as well as the symmetric method is negative, i.e., the estimates are low biased.

Compared to Table 4, Panel A of Table 5 shows that when using out-of-sample pricing, the estimated prices with the regular method improve significantly when the number of simulated paths, N , increases. The estimates from the symmetric method, however, are much less affected by the number of paths used. The reason for this is related to the over fitting and large variance in the

¹¹ Again, for all other values of the number of simulated paths with this number of regressors and when using other numbers of regressors with this number of simulated paths, the symmetric estimates are less biased.

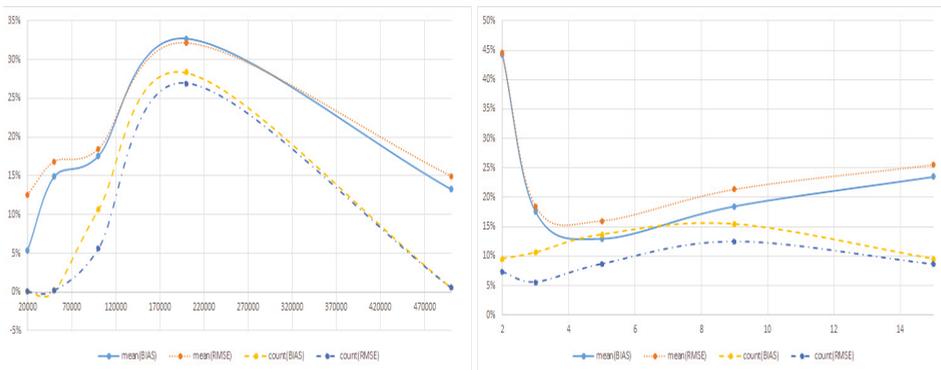
estimates in the cross-sectional regressions with the regular method, which for a given choice of L , becomes less of an issue with increasing N . When using symmetric pricing, this is much less of an issue since the regressors are bounded. Compared to Figure 3, Figure 4 shows that when using out-of-sample pricing, the relative performance of the symmetric method is much less dependent on the choice of N and L . In particular, the symmetric method now improves significantly on the regular method irrespective of the choice of L . Panel B of Table 5 shows that once $L = 3$ or more regressors are used, the RMSE of the symmetric method is around 20% of the RMSE of the regular method. In terms of the number of times the symmetric method leads to the smallest RMSE, this is around 90% or more for all values of L .

Table 5. Pricing errors across algorithm characteristics using out-of-sample pricing.

Panel A: Across Number of Simulated Paths N								
N	Absolute Metrics				Counting Metrics			
	Regular Call		Symmetric Call		Regular Call		Symmetric Call	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
20,000	-0.1286	0.1303	-0.0070	0.0163	0.9997	0.9997	0.0003	0.0003
50,000	-0.0974	0.0986	-0.0145	0.0166	0.9984	0.9978	0.0016	0.0022
100,000	-0.0837	0.0851	-0.0147	0.0157	0.9040	0.9472	0.0960	0.0528
200,000	-0.0565	0.0590	-0.0185	0.0190	0.7792	0.7878	0.2208	0.2122
500,000	-0.0706	0.0708	-0.0094	0.0106	0.9939	0.9942	0.0061	0.0058

Panel B: Across Number of Regressors L								
L	Absolute Metrics				Counting Metrics			
	Regular Call		Symmetric Call		Regular Call		Symmetric Call	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
2	-0.0974	0.0983	-0.0431	0.0438	0.9130	0.9322	0.0870	0.0678
3	-0.0837	0.0851	-0.0147	0.0157	0.9040	0.9472	0.0960	0.0528
5	-0.0488	0.0498	-0.0063	0.0079	0.8797	0.9203	0.1203	0.0797
9	-0.0418	0.0427	-0.0077	0.0091	0.8662	0.8890	0.1338	0.1110
15	-0.0450	0.0458	-0.0106	0.0117	0.9133	0.9210	0.0867	0.0790

This table shows pricing errors for the regular and symmetric method using out-of-sample pricing for various numbers of simulated paths N and number of regressors L . Results are based on $I = 100$ independent simulations. In each panel, we report results for the bias and RMSE in terms of the average metrics and counting metrics, i.e., the fraction of times a given method has the highest error metric.



(a) Relative performance across N using out-of-sample pricing **(b)** Relative performance across L using out-of-sample pricing

Figure 4. Relative pricing performance across algorithm characteristics using out-of-sample pricing. This figure plots the relative performance of the symmetric method compared to the regular method using out-of-sample pricing across the number of simulated paths, N , and number of regressors, L .

5.2. Extensions to Other Option Pricing Models

Until now, we have presented results for the simple Black–Scholes–Merton setup. The reason for this was obvious: we wanted to have fast and precise benchmark results available. Without these, it makes no sense to talk about one method being more efficient than another since we measure efficiency by loss functions such as the RMSE for which a benchmark is required. However, we have argued that since our results rely on nothing but simulation and regression, our conclusions should be valid for other settings in terms of asset dynamics and option payoffs for which PCS holds. We now consider two obvious alternatives and demonstrate that our previous conclusions indeed continue to hold. We first provide results when the option payoffs depend on the average of several assets in a multivariate model, and second, we consider the case where the asset dynamics are instead given by the stochastic volatility (SV) model of [Heston \(1993\)](#).

The case with options written on multiple assets is the most obvious generalization of the standard constant volatility case. Options can be written on the maximum, minimum or average of multiple assets. These types of payoff functions have been used widely in the literature, and the work in [Boyle and Tse \(1990\)](#) gave examples on where these types of options are traded. The work in [Stentoft \(2004a\)](#) demonstrated that as the dimension of the problem increases, simulation-based methods like the LSMC are the most efficient methods to use for pricing. While put-call symmetry properties have been established in several cases (see for example [Detemple \(2001\)](#)), the most clean-cut case occurs with options written on the geometric average, i.e., options for which the payoff is given by:

$$G = \max \left(0, \left(\prod_{l=1}^M S^m \right)^{\frac{1}{M}} - K \right), \tag{8}$$

where S^m , $m = 1, \dots, M$ are the prices of the underlying assets, M being the dimension, and K is the strike price as before. The reason that this case is a “clean-cut” example is that since the product of lognormals is lognormal, the pricing problem essentially reduces to that of pricing single asset options on an asset that follows a (particular and slightly non-standard) GBM.¹²

In the LSMC method, we again consider the complete set of polynomials of order $L = 3$ or less and therefore use a total of 10, 20 and 56 regressors when the dimension of the problem, M , is 2, 3 and 5, respectively. In all cases, we use $I = 100$ independent simulations with $N = 100,000$ simulated paths. We consider options with different features and different asset dynamics. In particular, we consider three values of the strike price, $K = [95, 100, 105]$, the time to maturity, $T = [0.5, 1, 2]$ years, the volatility, $\sigma = [10\%, 20\%, 40\%]$, the correlation between the assets, $\rho = [0.25, 0.50, 0.75]$, and the number of assets, $M = [2, 3, 5]$, for a total of $3^5 = 243$ options. Pricing options in multiple dimensions quickly become computationally complex, and for this reason, we consider a smaller sample of options than in the benchmark case. Overall, our results show that the average RMSE obtained with the symmetric method is 36% smaller than that obtained with the regular method, and the RMSE is smallest for 87% of the individual options when using the symmetric method. These results clearly show that symmetry is valuable also for more advanced multivariate models.

Figure 5a and Panel A of Table 6 show the results across the number of assets, M , and demonstrate clearly that our suggestion of pricing call options as put options becomes more important as the dimension, and hence the computational complexity, of the option increases. In particular, the relative error of the symmetric method in terms of RMSE in Figure 5a goes from 0.890–0.545 as the dimension increases from 2–5. Moreover, when the dimension of the problem is high, the symmetric method almost always, in 96.3% of the cases, has the lowest RMSE, as shown in Panel A of Table 6.

¹² The working version of this paper contains the full details on how to derive these dynamics.

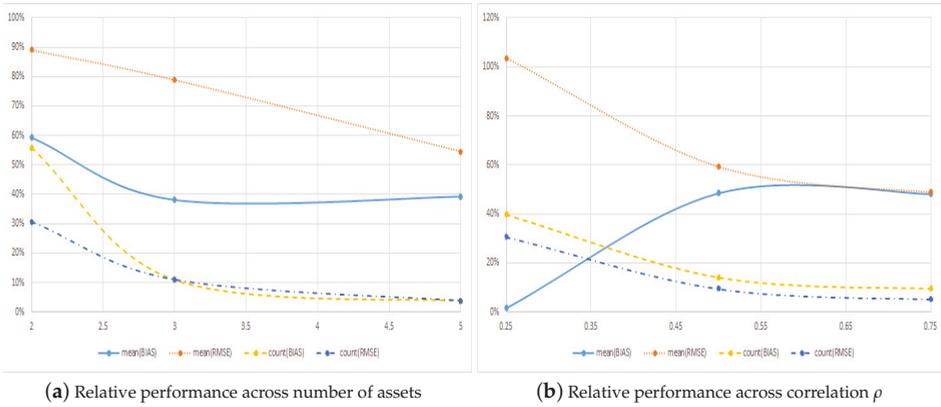


Figure 5. Relative pricing performance in a multivariate model. This figure plots the relative performance of the symmetric method compared to the regular method across the mean reversion rate κ , long-term variance, θ , initial volatility, V_0 , and correlation, ρ .

Table 6. Pricing errors across model characteristics in a multivariate model.

Panel A: Across Number of Assets M								
M	Absolute Metrics				Counting Metrics			
	Regular Call		Symmetric Call		Regular Call		Symmetric Call	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
2	0.0021	0.0054	0.0012	0.0048	0.6420	0.7654	0.3580	0.2346
3	0.0099	0.0123	0.0038	0.0097	0.9012	0.9012	0.0988	0.0988
5	0.0319	0.0329	0.0125	0.0180	0.9630	0.9630	0.0370	0.0370

Panel B: Across Correlation ρ								
ρ	Absolute Metrics				Counting Metrics			
	Regular Call		Symmetric Call		Regular Call		Symmetric Call	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
0.25	0.0073	0.0111	-0.0001	0.0115	0.7160	0.7654	0.2840	0.2346
0.5	0.0143	0.0161	0.0069	0.0096	0.8765	0.9136	0.1235	0.0864
0.75	0.0223	0.0234	0.0107	0.0114	0.9136	0.9506	0.0864	0.0494

This table shows pricing errors for the regular and symmetric method in a multivariate model for various values of the number of stocks M and correlation ρ . Results are based on $I = 100$ independent simulations with $N = 100,000$ paths. In each panel, we report results for the bias and RMSE in terms of the average metrics and counting metrics, i.e., the fraction of times a given method has the highest error metric.

Figure 5b and Panel B of Table 6 show the results across the correlation between the assets, ρ . From the figure, it is seen that symmetry becomes more important when correlations between assets increase. In particular, in terms of the absolute metrics, the regular method performs on par with the symmetric method when the correlation is low and $\rho = 0.25$, but when the correlation is high and $\rho = 0.75$, the error in pricing of the symmetric method is only around half that of the regular method, as shown in Figure 5b. Moreover, the symmetric method is always the method that yields the smallest errors for the largest fraction of the options. For example, when correlations are high among the underlying assets, the symmetric method has larger RMSE for only 4.94% of the options as shown in Panel B of Table 6.

The SV model of Heston (1993) is one of the most famous extensions to the constant volatility model. While it may not have been the first SV model, e.g., earlier examples include Hull and White (1987), Scott (1987) and Wiggins (1987), this particular model has emerged as the most important one and now serves as a benchmark against which many other SV models are compared. In the model of Heston (1993), the variance followed a Cox et al. (1985) process specified as:

$$dv_t = \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}dW_{2,t}^Q. \tag{9}$$

Here, κ represents the mean reversion rate of the variance, θ is the long-term variance, σ is the volatility of volatility (vol of vol) and the stock dynamics and variance are allowed to be correlated with correlation coefficient ρ . Put-call symmetry also holds in this model as demonstrated by, e.g., Battauz et al. (2014), which also listed the appropriate changes of parameters. In particular, the paper shows that in the model of Heston (1993), the following parity holds between a call option and a put option:

$$C(S, r, q, K, V_0, \theta, \kappa, \sigma, \rho) = P(K, q, r, S, V_0, \kappa\theta/(\kappa - \sigma\rho), \kappa - \sigma\rho, \sigma, -\rho). \tag{10}$$

In other words, a call option can be priced as a put option in which the stock price and the strike price and the interest rate and the dividend yield are interchanged, as was the case in the constant volatility setting, and in which the mean reversion, κ' , the long-term variance, θ' , and the correlation, ρ' , for the symmetric put option are changed to:

$$\kappa' = \kappa - \sigma\rho, \theta' = \theta\kappa/\kappa', \text{ and } \rho' = -\rho, \tag{11}$$

where κ , θ and ρ are the actual mean reversion, long-term variance and correlation.

We again price each of the individual options $I = 100$ -times with independently-simulated state variables used in the LSMC method, which we implemented using $N = 100,000$ paths. In the cross-sectional regressions, we use the complete set of polynomials in two dimensions of order $L = 3$ or less as regressors for a total of 10 regressors including the constant term. We consider three different values of the strike price, $K = [90, 100, 110]$, the time to maturity, $T = [0.5, 1, 2]$ years, the mean reversion rate, $\kappa = [1, 3, 5]$, the long-term variance, $\theta = [0.01, 0.04, 0.16]$, the initial level of the variance, $V_0 = [0.01, 0.04, 0.16]$, the volatility of volatility, $\sigma = [0.05, 0.1, 0.2]$, and the correlation between the two Brownian motions, $\rho = [-0.5, 0, 0.5]$, for a total of $3^7 = 2187$ options. Overall, our results show that the average RMSE obtained with the symmetric method is 14% smaller than that obtained with the regular method, and the RMSE is smallest for 61% of the individual options when using the symmetric method. These results clearly show that symmetry is valuable also for more advanced models, and though the results are somewhat closer to each other than with our benchmark model, this is to be expected since we are considering options here that have on average shorter maturities.

Figure 6 and Table 7 show the results across the various values of the mean reversion rate, κ , the long-term variance, θ , the initial level of the variance, V_0 , and the correlation, ρ . The first thing to notice from the table is that across all these interesting parameters and parameter values, the symmetric method most of the time outperforms the regular approach. The two exceptions to this are options with very low long-term variance and $\theta = 0.01$ and when the initial level of the variance is very high and $V_0 = 0.16$, where the RMSE is slightly lower with the regular method than with the symmetric method. In both of these cases, the symmetric method also has larger RMSE for more options than does the regular method, whereas in all the other cases, the symmetric method most often has the smallest RMSE.

In terms of performance across parameters, Figure 6 and Table 7 show that the symmetric method performs relatively better the faster the volatility mean reverts, i.e., the larger the value of κ in Figure 6a, and the larger the value of the long-term variance, i.e., the larger the value of θ in Figure 6b. The last of these findings is completely in line with the results from our benchmark model where symmetric pricing was most important for high values of the asset volatility. Figure 6d shows that the symmetric method performs relatively better when correlations are negative, which is indeed the empirically most relevant case in the stochastic volatility model of Heston (1993). When the correlation is highly negative and $\rho = -0.5$, the symmetric method has errors in terms of bias and RMSE that are around 25% lower than the regular method and has the largest errors for only around one third of the individual options. The effect on the relative performance of the initial variance, V_0 , is less clear cut though Figure 6c does indicate that the symmetric method performs relatively better when V_0 is not too extreme, and in particular not too high.

Table 7. Pricing errors across model characteristics in a stochastic volatility model.

Panel A: Across Mean Reversion κ									
κ	Absolute Metrics				Counting Metrics				
	Regular Call		Symmetric Call		Regular Call		Symmetric Call		
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	
1	0.0084	0.0139	0.0102	0.0128	0.5089	0.5638	0.4911	0.4362	
3	0.0055	0.0099	0.0066	0.0082	0.5446	0.6063	0.4554	0.3937	
5	0.0053	0.0093	0.0058	0.0074	0.5706	0.6516	0.4294	0.3484	

Panel B: Across Long-Term Variance θ									
θ	Absolute Metrics				Counting Metrics				
	Regular Call		Symmetric Call		Regular Call		Symmetric Call		
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	
0.01	0.0028	0.0077	0.0061	0.0081	0.4486	0.4938	0.5514	0.5062	
0.04	0.0045	0.0084	0.0057	0.0072	0.5748	0.6241	0.4252	0.3759	
0.16	0.0119	0.0170	0.0108	0.0130	0.6008	0.7037	0.3992	0.2963	

Panel C: Across Initial Variance V_0									
V_0	Absolute Metrics				Counting Metrics				
	Regular Call		Symmetric Call		Regular Call		Symmetric Call		
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	
0.01	0.0081	0.0115	0.0068	0.0090	0.6296	0.6818	0.3704	0.3182	
0.04	0.0035	0.0088	0.0035	0.0062	0.6337	0.7051	0.3663	0.2949	
0.16	0.0076	0.0128	0.0122	0.0131	0.3608	0.4348	0.6392	0.5652	

Panel D: Across Correlation ρ									
ρ	Absolute Metrics				Counting Metrics				
	Regular Call		Symmetric Call		Regular Call		Symmetric Call		
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE	
-0.5	0.0096	0.0120	0.0074	0.0092	0.6488	0.6927	0.3512	0.3073	
0	0.0060	0.0105	0.0070	0.0092	0.5075	0.6008	0.4925	0.3992	
0.5	0.0036	0.0105	0.0081	0.0100	0.4678	0.5281	0.5322	0.4719	

This table shows pricing errors for the regular and symmetric method in a stochastic volatility model for various values of the mean reversion rate, κ , long-term variance, θ , initial volatility, V_0 , and correlation, ρ . Results are based on $l = 100$ independent simulations with $N = 100,000$ paths. In each panel, we report results for the bias and RMSE in terms of the average metrics and counting metrics, i.e., the fraction of times a given method has the highest error metric.

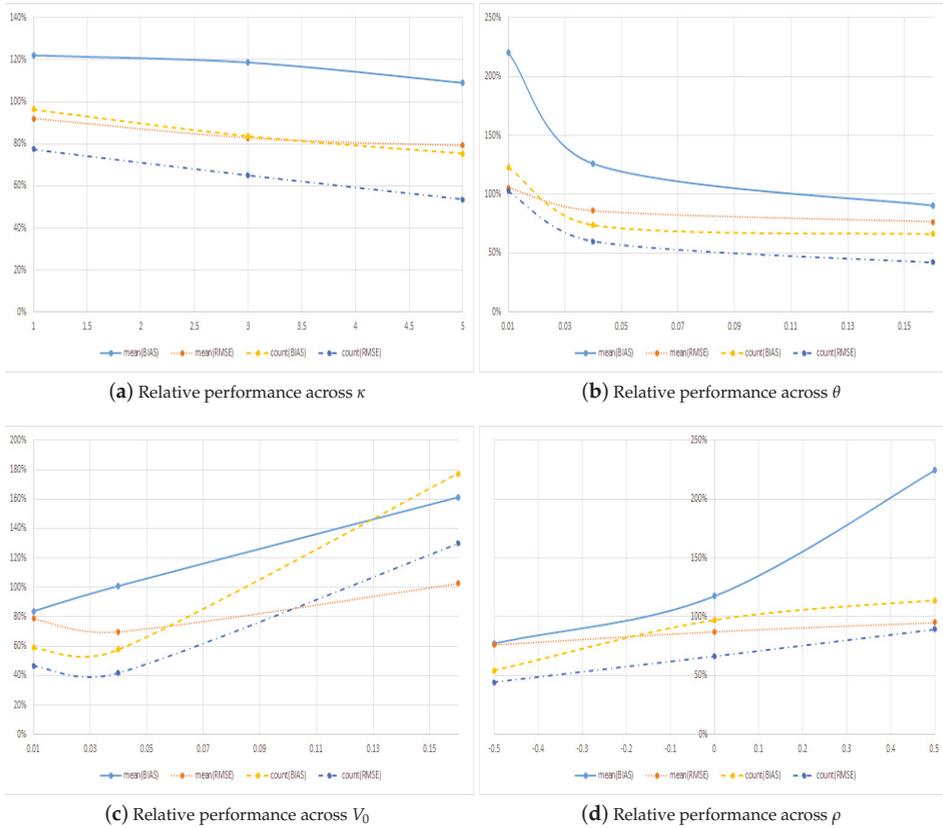


Figure 6. Relative pricing performance in a stochastic volatility model. This figure plots the relative performance of the symmetric method compared to the regular method across the mean reversion rate κ , long-term variance, θ , initial volatility, V_0 , and correlation, ρ .

6. Discussion

The fact that pricing call options using the symmetry method works best for most and along some dimensions almost all of the options considered is great news. However, since it does not perform the best for all the options, it leaves the obvious question of when to choose one method over the other. As it is, the only solid recommendations that arise from Sections 4 and 5 are that using the symmetric method with standard choices of the number of paths and number of regressors used in the LSMC method is relatively better the longer the maturity and the larger the volatility and that the methods become more similar when simulating a very large number of paths, e.g., when N is as large as 500,000, and that they diverge when using a large number of regressors, e.g., when L is as large as 15. In this section, we first examine the performance of the individual methods in terms of a relative efficiency measure, which compares the performance of a method to what could have been obtained optimally. Then, using properties of the out-of-sample method for pricing, we propose a method for selecting which specification to use across the methods and the number of regressors, which is simple to implement and achieves a very high degree of efficiency.

6.1. Efficiency as an Alternative Metric

Until now, we have compared performance metrics, i.e., the RMSE or the number of times a method works the best or worst, for the regular and symmetric pricing methods, respectively. An alternative and perhaps more interesting metric for “practitioners” is what one stands to lose in terms of increased pricing errors by picking and sticking to one particular method instead of using the optimal method for a given individual option in our sample. To examine this, we now consider a metric, which we will refer to as the “efficiency” given by the ratio of a specific method’s RMSE to the optimal and infeasible RMSE that could be obtained if one knew which method to use for each of the individual options.

Table 8 shows the efficiency of the two methods using in-sample pricing for various values of N and L in Columns 5 and 6. For comparison, the fraction of the options for which a particular model performs the best in terms of having the lowest RMSE is also reported in Columns 7 and 8.¹³ Panel A of Table 8 clearly shows that the symmetric method performs extremely well across the number of simulated paths, and one would never lose more than 7% from using this method. In fact, for most realistic specifications, i.e., when $N \geq 100,000$, the loss is less than 2%. The regular method, on the other hand, often has an efficiency of just around 20%, meaning that if this method was used to price the sample of options, one would lose around 80% compared to what could optimally be obtained.

Table 8. Efficiency across the number of paths N and regressors L .

Panel A: Results across Number of Paths N							
N	Call	RMSE		Efficiency		Count	
		S _{Call}	Optimal	Call	S _{Call}	Call	S _{Call}
20,000	0.0426	0.0178	0.0166	38.91%	92.98%	30.85%	69.15%
50,000	0.0480	0.0131	0.0125	26.00%	95.56%	20.00%	80.00%
100,000	0.0602	0.0104	0.0102	16.99%	98.33%	11.81%	88.19%
200,000	0.0540	0.0100	0.0100	18.49%	99.95%	2.34%	97.66%
500,000	0.0556	0.0124	0.0123	22.04%	98.83%	11.65%	88.35%

Panel B: Results across Number of Regressors L							
L	Call	RMSE		Efficiency		Count	
		S _{Call}	Optimal	Call	S _{Call}	Call	S _{Call}
2	0.0832	0.0351	0.0345	41.41%	98.17%	10.14%	89.86%
3	0.0602	0.0104	0.0102	16.99%	98.33%	11.81%	88.19%
5	0.0201	0.0117	0.0099	49.34%	84.48%	60.99%	39.01%
9	0.0328	0.0154	0.0150	45.78%	97.21%	28.77%	71.23%
15	0.0625	0.0187	0.0184	29.39%	98.00%	24.45%	75.55%

This table shows the efficiency across the number of paths, N , and regressors, L . Efficiency is measured as the optimal RMSE, conditional on knowing which of the two methods yields the lowest RMSE, as a fraction of the RMSE of the regular or symmetric method, respectively. Panel A reports results for different values of the number of simulated paths, N , and Panel B reports results for different values of the number of regressors, L . In addition to the efficiency, the table also reports the fraction of times for which the regular and symmetric method provide the lowest RMSE, respectively.

Panel B of the table is, given the results in the previous section on robustness, even more interesting. In particular, the previous results showed that for some specification, i.e., when picking $L = 5$, the symmetric method actually has larger RMSE than the regular method for most options. The row labelled $L = 5$ in Table 8, however, shows that even in this case where the symmetric RMSE is the lowest for only 39% of the options, the method’s efficiency is above 84%. That is, even for settings when the regular method is the best, measured by minimizing the RMSE, for 61% of the options when

¹³ These numbers are the “inverse” of the counting metrics used in previous tables.

using the symmetric method, you would not lose more than 16% compared to what could be optimally obtained had you known what would be the best method to use for the individual options. It is also striking that if you, on the other hand, would use the regular method for all options, the efficiency is only around 49% in spite of the fact that this is the method that has the lowest RMSE for most of the options.

Table 9 shows the efficiency of the two methods using out-of-sample pricing for various values of N and L in Columns 5 and 6. The first thing to notice from this table is that when using out-of-sample pricing, i.e., when a new set of paths is used for pricing, the efficiency of the symmetric method is extremely high, and often above 99%, across both the choice of the number of paths, N , and the number of regressors, L . Compared to the in-sample results in Table 8, the efficiency of the symmetric method is most of the time improved, the exception being when using $N = 200,000$ paths in the simulation. For the regular method, on the other hand, efficiency is generally much lower, as low as 15%, and does not improve in any systematic way when using out-of-sample pricing. In conclusion, although the symmetric method is not always the model that has the smallest RMSE, the efficiency of this method is generally very high, always significantly higher than that of the corresponding regular method, and therefore, the costs of using this method are always reasonably low. Our suggestion is therefore very naturally to use the symmetric method for call option pricing.

Table 9. Efficiency across number of paths N and regressors L using OSpricing.

Panel A: Results across Number of Paths N							
N	RMSE			Efficiency		Count	
	Call	SCall	Optimal	Call	SCall	Call	SCall
20,000	0.1303	0.0163	0.0163	12.54%	100.00%	0.03%	99.97%
50,000	0.0986	0.0166	0.0166	16.81%	100.00%	0.22%	99.78%
100,000	0.0851	0.0157	0.0156	18.38%	99.86%	5.28%	94.72%
200,000	0.0590	0.0190	0.0185	31.36%	97.52%	21.22%	78.78%
500,000	0.0708	0.0106	0.0106	14.93%	99.99%	0.58%	99.42%

Panel B: Results across Number of Regressors L							
L	RMSE			Efficiency		Count	
	Call	SCall	Optimal	Call	SCall	Call	SCall
2	0.0983	0.0438	0.0437	44.42%	99.63%	6.78%	93.22%
3	0.0851	0.0157	0.0156	18.38%	99.86%	5.28%	94.72%
5	0.0498	0.0079	0.0079	15.79%	99.03%	7.97%	92.03%
9	0.0427	0.0091	0.0090	21.01%	98.49%	11.10%	88.90%
15	0.0458	0.0117	0.0116	25.36%	99.36%	7.90%	92.10%

This table shows the efficiency across the number of paths, N , and regressors, L , using out-of-sample pricing. Efficiency is measured as the optimal RMSE, conditional on knowing which of the two methods yields the lowest RMSE, as a fraction of the RMSE of the regular or symmetric method, respectively. Panel A reports results for different values of the number of simulated paths, N , and Panel B reports results for different values of the number of regressors, L . In addition to the efficiency, we also report the fraction of times for which the regular and symmetric method provide the lowest RMSE, respectively.

6.2. Picking the Best Configuration

Although we recommend to always use the symmetric method for call option pricing, you may still wonder if it is possible to improve on this recommendation, i.e., if it is possible to pick the “right” model using some “observables”. This is essentially a question of classification. A straightforward classification variable is the estimated price. In particular, we know that when using the out-of-sample pricing technique, estimates are in expectation low biased. Moreover, while we expect the estimates to increase when increasing the number of regressors, L , initially, as this improves the polynomial approximation, when L becomes very large and over fitting to the paths used to determine the optimal exercise strategy becomes a problem, the estimated out-of-sample price could decrease. When comparing results for several different values of L and different methods, i.e., regular versus symmetric,

one could therefore propose to choose the method that maximizes the out-of-sample price. In particular, this should result in picking the method that has the smallest bias, and this would potentially also be the method with a small RMSE. The results from implementing this classification strategy are shown in Table 10.

Table 10. Efficiency across the number of regressors L using OS pricing.

Panel A: Results for Individual Values of L										
L	RMSE			Classified		Local Efficiency		Global Efficiency		
	Call	SCall	Optimal	RMSE	Picked	Efficiency	Call	SCall	Call	
2	0.0983	0.0438	0.0437	0.0437	93.34%	99.83%	44.42%	99.63%	7.45%	16.72%
3	0.0851	0.0157	0.0156	0.0157	95.36%	99.65%	18.38%	99.86%	8.61%	46.79%
5	0.0498	0.0079	0.0079	0.0079	95.74%	99.66%	15.79%	99.03%	14.72%	92.27%
7	0.0440	0.0082	0.0081	0.0081	96.77%	99.78%	18.36%	98.31%	16.65%	89.17%
9	0.0427	0.0091	0.0090	0.0090	97.44%	99.90%	21.01%	98.49%	17.15%	80.41%
11	0.0430	0.0100	0.0099	0.0099	97.89%	99.96%	22.92%	98.85%	17.02%	73.40%
15	0.0458	0.0117	0.0116	0.0116	99.17%	99.99%	25.36%	99.36%	16.01%	62.75%

Panel B: Results for Multiple Values of L										
L	RMSE			Classified		Local Efficiency		Global Efficiency		
	Call	SCall	Optimal	RMSE	Picked	Efficiency	Call	SCall	Call	
All	0.0379	0.0074	0.0073	0.0074	90.43%	98.98%	99.99%	99.99%	19.33%	98.36%

This table shows the relative efficiency of the two methods using out-of-sample pricing across the number of regressors, L . Efficiency is measured as the optimal RMSE, conditional on knowing which of the methods yields the lowest RMSE for a given individual option in our large sample of 3125 options, as a fraction of the RMSE of the regular and symmetric method, respectively. Panel A report results for different values of the number of regressors used in the cross-section regression, L , and Panel B reports results across all the values of L in Panel A. The result in the column headed “Optimal” corresponds to selecting the method with the minimum RMSE either for a given L , in Panel A, or across all values of L , in Panel B. The results in the column headed “Classified” correspond to what would be obtained if the method with the highest price is used, either for a given L or across all values of L and reports the resulting RMSE, the fraction of options for which the method with the lowest RMSE was actually picked and the efficiency of this method compared to the corresponding optimal method. In Panel A, “Local Efficiency” is measured relative to the optimal RMSE for a given value of L . In Panel B, “Local Efficiency” is measured relative to the minimum RMSE obtained for the regular and symmetric method, respectively, across all values of L . In both panels, results in the column headed “Global Efficiency” report the efficiency of the two methods compared to the best possible RMSE of 0.0073 reported in Panel B.

Panel A in Table 10 reports results for individual values of L , i.e., when the method, regular or symmetric, that has the highest price for a given value of L is picked. The first thing to notice from this panel is that the right method for a given option is picked at least 93% of the time, and the efficiency of this method is always above 99%. The symmetric method clearly performs the best on average for all values of L , and this method does have a very high “local” efficiency, that is compared to the optimal RMSE for a particular value of L . The regular method, on the other hand, has a much lower efficiency. Compared to the efficiency of the individual methods, the panel shows that classification according to maximum price does improve on the RMSE in all but one case. In terms of “global” efficiency though, the performance of the methods varies greatly across L .

Panel B in Table 10 reports results across all the values of L used in Panel A, i.e., in this panel, the optimal RMSE is picked across both methods and the values of L . The first thing to notice from this panel is that the classification method performs very well and picks the right method more than 90% of the time and has a very high efficiency of close to 99%. Picking the symmetric method that has the highest price across L also results in estimates that are very efficient, although the optimal RMSE is slightly lower. The efficiency of the regular method is below 20% when compared to the globally optimal method, although measured locally across L picking the method with the highest price results in estimates with an RMSE very close to the minimum RMSE for this method.

The results above show that when using out-of-sample pricing, it is possible to derive a simple classification algorithm that achieves very high efficiency. In reality, the algorithm ends up picking the

symmetric method most of the time, and if you only pick within this method, the loss in efficiency is very small, i.e., it decreases from 98.98–98.36%. Thus, it is possible to save on the computational time by only considering this method. Moreover, for most of the options, the highest price is achieved with $L = 5$ when using the symmetric method, and if, instead of picking among all the possible values of L in the table, you only consider $L \leq 7$, the global efficiency decreases only marginally to 98.24%.¹⁴ This approach thus yields very good estimates across our large sample of options, and it is easy to implement. It is not possible to come up with a similar approach when using in-sample pricing though.

7. Conclusions

This paper shows that it is possible to improve significantly on the estimated call prices obtained with a state-of-the-art algorithm, the Least Squares Monte Carlo (LSMC) method of Longstaff and Schwartz (2001), for option pricing using regression and Monte Carlo simulation by using Put-Call Symmetry (PCS). PCS holds widely and in the classical Black–Scholes–Merton case, for example, implies that a call option has the same price as an otherwise similar put option where strike price and stock level and where interest rate and dividend yield are interchanged, respectively. The immediate implication is that path-wise payoffs in simulations are bounded above by the strike price instead of being unbounded, and for methods that use regression to determine the optimal early exercise strategy, this leads to much improved estimates of the stopping time and more precise option price estimates as measured by, for example, the Root Mean Squared Error (RMSE). Our results show that, for a large sample of options with characteristics of relevance in real-life applications, the symmetric method on average performs much better than the regular pricing method, is the best method for most of the options, never performs very poorly and as a result is very efficient compared to an optimal, but unfeasible method that picks the method with the smallest RMSE. When using out-of-sample pricing, a simple classification algorithm is proposed that, by optimally selecting among estimates from the symmetric method with a reasonably small order used in the polynomial approximation, achieves a relative efficiency of more than 98% compared to the infeasible, but optimal method that minimizes the RMSE across all estimates. Our results also show that the relative importance of using the symmetric method increases with option maturity and with asset volatility and using symmetry methods to price long-term options in high volatility situations improves massively on the price estimates. The LSMC method is routinely used to price real options, many of which are call options with long maturity on volatile assets, for example energy. We therefore conjecture that pricing such options using the symmetric method could improve the estimates significantly by decreasing their bias and RMSE by orders of magnitude.

Funding: This research received no external funding.

Acknowledgments: I thank Pascal Francois for bringing the put-call symmetry property to my attention and the three reviewers for excellent comments that led to a much improved paper. Dayi Li provided expert research assistance. For additional numerical results and further details, please consult the online working paper version, which is available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3362426. More than 5,000,000 artificial individual options were priced in this paper, which would have been impossible without the computational resources made available by the Canadian Foundation for Innovation (CFI).

Conflicts of Interest: The author declares no conflict of interest.

References

- Battauz, Anna, Marzia De Donno, and Alessandro Sbuelz. 2014. The Put-Call Symmetry for American Options in the Heston Stochastic Volatility Model. *Mathematical Finance Letters* 7: 1–8.
- Boyle, Phelim P., and Yiu Kuen Tse. 1990. An Algorithm for Computing Values of Options on the Maximum or Minimum of Several Assets. *Journal of Financial and Quantitative Analysis* 25: 215–27. [CrossRef]

¹⁴ The mode of the number of regressors, L , for the regular method is seven, on the other hand.

- Chow, Yuan-Shih, Herbert Robbins, and David Siegmund. 1971. *Great Expectations: The Theory of Optimal Stopping*. New York: Houghton Mifflin.
- Cox, John C., Jonathan E. Ingersoll, and Stephen A. Ross. 1985. A Theory of the Term Structure of Interest Rates. *Econometrica* 53: 385–408. [\[CrossRef\]](#)
- Cox, John C., Stephen A. Ross, and Mark Rubinstein. 1979. Option pricing: A simplified approach. *Journal of Financial Economics* 7: 229–63. [\[CrossRef\]](#)
- Detemple, Jerome. 2001. American Options: Symmetry Properties. In *Handbooks in Mathematical Finance: Topics in Option Pricing, Interest Rates and Risk Management*. Edited by Elyes Jouini, Jaksa Cvitanic and Marek Musiela. Cambridge: Cambridge University Press, pp. 67–104.
- Duffie, Darrell. 1996. *Dynamic Asset Pricing Theory*. Princeton: Princeton University Press.
- Grabbe, J. Orlin. 1983. The Pricing of Call and Put Options of Foreign Exchange. *Journal of International Money and Finance* 2: 239–53. [\[CrossRef\]](#)
- Heston, Steven L. 1993. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *Review of Financial Studies* 6: 327–43. [\[CrossRef\]](#)
- Hull, John, and Alan White. 1987. The Pricing of Options on Assets with Stochastic Volatilities. *Journal of Finance* 42: 281–300. [\[CrossRef\]](#)
- Karatzas, Ioannis. 1988. On the Pricing of American Options. *Applied Mathematics and Optimization* 17: 37–60. [\[CrossRef\]](#)
- Longstaff, Francis A., and Eduardo S. Schwartz. 2001. Valuing American Options by Simulation: A Simple Least-Squares Approach. *Review of Financial Studies* 14: 113–47. [\[CrossRef\]](#)
- McDonald, Robert C., and Mark Schroder. 1998. A Parity Result for American Options. *Journal of Computational Finance* 1: 5–13. [\[CrossRef\]](#)
- Royden, Halsey. 1988. *Real Analysis*. Upper Saddle River: Prentice Hall, Inc.
- Schroder, Mark. 1999. Changes of Numeraire for Pricing Futures, Forwards, and Options. *Review of Financial Studies* 12: 1143–63. [\[CrossRef\]](#)
- Scott, Louis O. 1987. Option Pricing When the Variance Changes Randomly: Theory, Estimation, and an Application. *Journal of Financial and Quantitative Analysis* 22: 419–38. [\[CrossRef\]](#)
- Stentoft, Lars. 2004a. Assessing the Least Squares Monte-Carlo Approach to American Option Valuation. *Review of Derivatives Research* 7: 129–68. [\[CrossRef\]](#)
- Stentoft, Lars. 2004b. Convergence of the Least Squares Monte Carlo Approach to American Option Valuation. *Management Science* 50: 1193–203. [\[CrossRef\]](#)
- Wiggins, James B. 1987. Option Values Under Stochastic Volatility: Theory and Empirical Estimates. *Journal of Financial Economics* 19: 351–72. [\[CrossRef\]](#)



© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Article

Positive Liquidity Spillovers from Sovereign Bond-Backed Securities

Peter G. Dunne

Central Bank of Ireland, Dublin 1 D01 F7X3, Ireland; peter.dunne@centralbank.ie

Received: 10 March 2019; Accepted: 3 April 2019; Published: 9 April 2019

Abstract: This paper contributes to the debate concerning the benefits and disadvantages of introducing a European Sovereign Bond-Backed Securitisation (SBBS) to address the need for a common safe asset that would break destabilising bank-sovereign linkages. The analysis focuses on assessing the effectiveness of hedges incurred while making markets in individual euro area sovereign bonds by taking offsetting positions in one or more of the SBBS tranches. Tranche yields are estimated using a simulation approach. This involves the generation of sovereign defaults and allocation of the combined credit risk premium of all the sovereigns, at the end of each day, to the SBBS tranches according to the seniority of claims under the proposed securitisation. Optimal hedging with SBBS is found to reduce risk exposures substantially in normal market conditions. In volatile conditions, hedging is not very effective but leaves dealers exposed to mostly idiosyncratic risks. These remaining risks largely disappear if dealers are diversified in providing liquidity across country-specific secondary markets and SBBS tranches. Hedging each of the long positions in a portfolio of individual sovereigns results in a risk exposure as low as that borne by holding the safest individual sovereign bond (the Bund).

Keywords: safe assets; securitisation; dealer behaviour; liquidity; bid–ask spread

1. Introduction

Considerable ambiguity exists regarding the likely liquidity effects of introducing Sovereign Bond-Backed Securities, as proposed by Brunnermeier et al. (2017). The competing forces are (i) a reduced supply of the underlying asset and (ii) improved prospects for the more efficient management of inventory risks. This paper argues that trading costs in country-specific European sovereign bond markets would be constrained by an arbitrage relation as a result of the existence of liquid Sovereign Bond-Backed Securities (SBBS) and this would dominate any negative supply effects.¹ This outcome relies on sufficient diversification and hedging by dealers in a truly European-wide context. In this sense, and in some other respects to be discussed below, the outcome envisaged is a deeply integrated market in which SBBS become the primary focus for price discovery while trading costs in most national markets shrink to within tight bounds of those in SBBS markets.

Brunnermeier et al. (2016) and Brunnermeier et al. (2017) propose the issuance of sovereign bond-backed securities (SBBS) in the euro area with tranches that would be sequentially exposed to losses arising from defaults on any of the underlying individual sovereign securities. The senior tranche would make up the majority of the securitisation and be extremely low risk. The junior tranches would be more sensitive to early signs of defaults, but, at the same time, maintain some of the risk reducing benefits of diversification. In theory, this proposal has the potential to induce bond holders to diversify beyond their national sovereign markets and this could substantially weaken the sovereign-bank

¹ A wider range of potential effects are considered in the report by the ESRB High-Level Task Force on Safe Assets (2018).

doom-loop that was prevalent during the Sovereign Debt Crisis in Europe. The question remains whether it would work in practice.

Traditionally, banks hold sovereign bonds as collateral for use in short-term treasury management and in official monetary policy operations. This is encouraged by the fact that such assets are zero risk weighted for capital adequacy purposes. During the Great Recession and euro area Sovereign Debt Crisis, many European banks experienced widespread defaults on their loan portfolios and they required recapitalisation from their governments. Sovereigns simultaneously experienced severe current account deficits and expected imminent high bank bail-out costs which drove their yields to record levels. The higher risk of sovereign default fed back into credit and counter-party risks for the linked banks and this exacerbated the expected bail-out costs for sovereigns. These linked risk exposures generated what was referred to as a “vicious circle” at a euro area summit of heads of governments in 2012. It was also referred to as the “doom loop” by [Farhi and Tirole \(2018\)](#) or as the “diabolic loop” by [Brunnermeier et al. \(2016\)](#).

Proposals to break the bank-sovereign vicious circle have been debated as part of a two-year reassessment by the Basel Committee on Banking Supervision of the Regulatory Treatment of Sovereign Exposures (RTSE) (see [Basel Committee on Banking Supervision 2017](#)). The Committee decided to retain zero risk weights on domestic sovereign exposures acknowledging that there was no consensus among supervisors, experts, and economists on alternative policy options to amend the framework. Alternative approaches to breaking the doom loop have therefore appeared and the Sovereign Bond Backed Securitisation proposal by [Brunnermeier et al. \(2016\)](#) is just one of these. Several related proposals are assessed by [Leandro and Zettelmeyer \(2018\)](#) and the analysis below would have some relevance for the ‘Blue bond’ proposal discussed there since that proposal gives rise to a similarly liquid low risk bond that could be used in hedging by dealers.

An apparent drawback of the SBBS proposal is that it would leave already small and illiquid sovereign bond markets with reduced free float and this would do further harm to liquidity and raise issuance costs. If orders arrive infrequently in smaller sovereign bond markets—and if orders of a type needed to reduce inventories are inelastic with respect to dealer pricing—then inventory positions will be held for longer and providing continuous liquidity would be costly. It is frequently argued that such costs would rise if the national free float is reduced by absorption into a securitisation.

The analysis below suggests that such a conclusion arises only when national markets are considered in isolation. A wider perspective introduces the potential for liquidity spillovers due to hedging and diversification in market making. Rather than reducing liquidity, the effects of SBBS on sovereign bond markets should have more in common with how opportunities to trade ‘To-Be-Announced’ US Agency Mortgage Backed Securities contracts (TBA-MBS) affect trading costs in Specific-Pools of Mortgage Backed Securities, even those that are not cheapest to deliver.²

If SBBS are very liquid, they present dealers with instantaneous access to offsetting positions following trades in any given sovereign market and this would establish a valid arbitrage relation to constrain trading costs in those markets. SBBS markets (particularly for the senior tranche) are likely to be deeper and more liquid than smaller euro area sovereign bond markets due mainly to the greater amounts of SBBS in issue relative to what is typical for any individual sovereign. However, this is also thought likely because SBBS markets (having factor-like properties) would acquire benchmark status. [Yuan \(2005\)](#) shows that the presence of benchmarks gives rise to liquidity in related markets.³

² [Gao et al. \(2017\)](#) describe how, in that market, dealers typically hedge inventory risk in their Specific-Pool exposures with offsetting TBA trades and they show that impediments to hedging can reduce such liquidity. More interestingly, they conclude that the presence of TBA markets has very widespread beneficial effects on liquidity significantly beyond the mortgage pools that are cheapest to deliver. This is also traced to the ability to hedge inventory holding risk.

³ In a related paper, the acquisition of benchmark status in pre-crisis European sovereign bond markets is examined in [Dunne et al. \(2007\)](#). Benchmarks tend to become liquid as they are the location for discovery of the systematic component of the risk premium (in this case, it is envisaged that the different tranches of SBBS would be benchmarks for credit risks within different categories of the market).

An important difference between the SBBS proposal and the situation discussed in Yuan (2005) is that increased supply of SBBS directly implies a reduction in the supply of individual sovereign bonds outstanding. Yuan assumes, in contrast, that benchmarks enhance the likelihood of issuance of corporate bonds and more issuance improves liquidity. However, a positive externality that Yuan's analysis does not take into account is the hedging benefits that dealers can avail of in the presence of liquid benchmarks. It also leaves out post-hedging diversification benefits. These externalities are worthy of further examination.

Although we lack actual evidence of the characteristics of SBBS, it is possible to assess their likely behaviour (or how they would have behaved in the past) through a simulation approach. The present analysis derives SBBS yield estimates using the simulation approach of Schönbucher (2003). This permits an assessment of hedge effectiveness and it shows that dealer inventory positions in most national markets can be, to some extent, hedged by using one or more of the tranches of the sovereign bond-backed securitisation. More importantly, it turns out that what cannot be hedged is largely idiosyncratic and diversifiable by dealers who are active in many markets. Since inventory risk only requires compensation for its systematic component, the liquidity benefits of SBBS are enhanced by diversification.

Hedge effectiveness is assessed by measuring the risk of the optimally-hedged position relative to the unhedged portfolio, as in Bessler et al. (2016). The additional diversification benefits are assessed by assuming that dealers typically have positions in all sovereigns (with both equal weights on individual sovereigns and market-size based weights). A core result from the analysis is that trading costs in SBBS markets would determine limits on the size of trading costs in national sovereign bonds.⁴

The following section outlines the type of hedging behaviour by dealers that is deemed likely when sovereign bond-backed securities exist. The relevant literature is then surveyed. Methodologies used to estimate SBBS yields, to measure time-varying hedge ratios and to assess diversification benefits, are then explained. The discussion of results from the application of these methods follows. The conclusion gives some indication of overall liquidity benefits that would flow from the existence of SBBS based on the recent history.

2. Hedging, Arbitrage and Diversification

The benefit of hedging with a highly-liquid, contemporaneously correlated, asset is clear-cut in the extreme case of perfect correlation. Let ask and bid prices in the bond market be denoted (**a**) and (**b**) and those in the SBBS market be (**A**) and (**B**), respectively. Assuming no frictions (i.e., no basis, coordination, execution or timing risks and no variability in market making risks or risk aversion), then arbitrage and competition between dealers should keep the two bid–ask spreads close to each other.⁵ Perfect correlation in the underlying values of the two securities and the assumption of instantaneous availability of trading opportunities in the highly liquid asset allow us to subtract the common underlying value changes, $V(t)$, from all bid and ask prices in each period t , leaving a^* , A^* , b^* and B^* as timeless (where starred variables are deviations from the relevant common $V(t)$).

A dealer who acquires a long position at a price (**b**) can immediately sell an equal amount in the SBBS market at price (**B**). This leaves the position hedged against movements in V until the bond is sold again at a price (**a**) and the SBBS is simultaneously bought at (**A**). Regardless of common movements in V , there is a profit for the dealer of $B^* - b^* + a^* - A^*$. This profit is trivially increasing in the difference between the spreads $s-S$. In a competitive market, we would expect such differences in spreads to be competed away (excluding any extra costs associated with operating in the more

⁴ Whether these bounds are sufficient to improve on current trading costs is moot. Even if the costs of hedging with SBBS were to exceed current trading costs in national markets, their use in this way would still be relevant in minimising the extent of any deterioration in trading costs due directly to reductions in the free float as a result of the securitisation.

⁵ This also relies, for simplicity, on the assumption that there is symmetry in the positioning of spreads relative to the underlying value. If not, then the proposition that follows applies on average across many trades.

general environment). The spread in the bond market will, in this case, be primarily determined by the required bid–ask spread in the SBBS market.

If there is only a partial correlation between the value of the SBBS and that of the bond then the remaining risk must be managed somehow. In this case, a diversification strategy should be very effective since remaining risks would largely be idiosyncratic. The finding of a substantial reduction in risk due to diversification can be regarded as indirect evidence that SBBS behave like systematic risk factors.

Since only the non-diversifiable component of the unhedged risk will require compensation, the link between the size of the spread in the hedge instrument and that in the asset being hedged is relatively unaffected relative to the case of a perfectly correlated hedge. Standard diversification arithmetic implies that, increasing the extent of diversification from one sovereign market to 11 uncorrelated markets with similar-sized risks and activity, would produce more than a 70% reduction in risk exposure. Hence, if dealers are active in providing liquidity across all European sovereign markets that contribute to the securitisation, then any unhedged risks will largely average-out in their trading portfolios.⁶

It is important to draw a distinction between hedging with offsetting positions in an asset possessing highly correlated contemporaneous price movements and the alternative of hedging with a futures contract on the underlying. Futures contracts involve risk premiums that impose direct cost on dealers. There is also a considerable difficulty in correctly matching the expected duration of inventory holding with the expiry time of the associated futures contract.⁷ Matching the size of an inventory position with standardized sizes in the futures market may also be an issue. Hedging with CDS contracts is similar to the use of a correlated asset but this analogy is also incomplete. The CDS premium will only be highly correlated with the credit risk of the underlying security. The disadvantage of the CDS when compared with a factor-like hedge is that it does not hedge systematic interest rate risk. CDS trading also tends to be less liquid than the underlying.

3. Microstructure Literature

Standard models of market making behaviour assume that bid and offer quotes involve a fixed component to cover order processing costs, an inventory management component (e.g., [Amihud and Mendelson 1980](#); [Demsetz 1968](#); [Tinic and West 1972](#); [Ho and Stoll 1981](#); [Ho and Stoll 1983](#) and [Stoll 1978](#)) and an adverse selection component to protect the dealer from losses to more informed traders (e.g., [Copeland and Galai 1983](#); [Easley and O'Hara 1987](#) and [Glosten and Milgrom 1985](#)). In the case of sovereign bonds, however, the adverse selection component is regarded as small since nearly all information about underlying value of such bonds is public.⁸

In respect of the inventory imbalance component, [Stoll \(1978\)](#) was one of the first to empirically assess the differential effects of systematic and idiosyncratic risk on dealer pricing. [Huang and Stoll \(1997\)](#) simply assert that the inventory component is mainly related to dealers' portfolio-wide inventory imbalances rather than stock-specific imbalances. They use this fact to separate the inventory component of the spread from the adverse selection component. More recently, [Hagströmer et al. \(2016\)](#) develops a structural model for price formation and liquidity supply explaining how inventory

⁶ It may also be supposed that this benign outcome would be compromised if the SBBS has a difficult-to-forecast correlation with the bond (i.e., if out of sample hedge ratios turn out to be less efficient than they could have been). This is really a type of operational risk and (assuming forecasts are as efficient as possible ex ante) this also gives rise to mostly idiosyncratic and diversifiable risks.

⁷ [Bessler et al. \(2016\)](#) point out that several futures markets for individual sovereign bonds existed pre-EMU and that the alignment of yields during the years of the Great Moderation was largely responsible for the disappearance of all but the German Bund futures. The Great Financial Crisis and the Sovereign Debt Crisis in Europe ultimately led to the reintroduction of futures on Italian BTPs and French OATs. Futures on Spanish Bonos only reappeared in 2015. [Naik and Yadav \(2003\)](#) examine the use of futures to hedge interest rate risk (undesirable duration) in sovereign bond portfolios of dealers and they find support for the propositions about hedging behaviour by [Froot and Stein \(1998\)](#).

⁸ For example, [Naik and Yadav \(2003\)](#) strongly reject the notion that dealers benefit from their information about orderflow even in the relatively concentrated UK Gilts market.

pressure and price discovery is influenced by joint trading of the same or a similar asset in a parallel market. This shows that where hedging opportunities exist trading costs are partially determined by characteristics of the hedge instrument rather than the specific asset.

Naik and Yadav (2003) provide extensive evidence of the use of futures by dealers in the UK Gilts market to hedge duration (they do not discuss credit risk hedging). They measure risk exposure in two ways; using value-weighted average duration of the whole portfolio of spot and futures bond positions of 15 dealers and alternatively, a value-weighted average of the duration-weighted hedge ratios (where hedge ratios are betas from regressions of returns on individual bonds on the long-gilt futures return). The two measures are very highly correlated (0.92). They show that dealers maintain a short position on average, but the spot and futures positions diverge in opposite directions from their means. They conclude that dealers actively use the futures contract to hedge their spot exposures in order to maintain a target level of interest rate exposure.

This hedging behaviour is consistent with theoretical predictions of Amihud and Mendelson (1980). The evidence of partial hedging supports the findings of Stulz (1996) and suggests that what is not hedged is idiosyncratic (only interest rate risk is being hedged in this case as the sample did not involve significant fluctuations in UK sovereign credit risk). The Naik and Yadav study goes on to show that hedging varies with the costs of hedging, volatility of holding returns, risk aversion and capital shortage. They also show that bigger dealing banks hedge less but do not seem to make profits from this—implying that their knowledge of a significant fraction of orderflow does not translate into an economic advantage.

4. Methodology

4.1. Derivation of SBBS Yields

Since sovereign bond-backed securities did not exist in the past we rely on estimates of yields on such instruments based on a simulation approach proposed by Schönbucher (2003)—this approach has already been implemented for the case of the SBBS proposal by De Sola Perea et al. (2017) and the ESRB High-Level Task Force on Safe Assets (2018). It is important to state that a motivating principle for the Schönbucher (2003) method is to retain the properties of the underlying relationship between yields and what they imply for default probabilities—including changing correlations and dynamic dependencies. Hence, the estimated SBBS yields in this case are not just some linear combination of the underlying securities (those used as backing for the securitisation). If that were the case, the linear relation would exactly determine the correlations that we rely on for hedge selection. The Schönbucher approach retains the variable probabilities of default in the underlying securities as well as their time-varying interdependencies.

The simulations that produce the SBBS yields are conducted on a daily basis over the period from soon after the introduction of the euro to the end of 2017. The simulations involve draws from 11 of the main sovereign bond yield-spreads (according to the report of the ESRB Task Force on Safe Assets, this group of sovereigns covers approximately 97.5% of outstanding sovereign debt in the euro area). The time period covered contains a lot of variability in circumstances including; the pre-2008 Great Moderation period, the period of the financial crisis in the wake of the Lehman Brothers default, the euro area sovereign debt crisis of 2009–2012 and the subsequent gradual improvement in euro area sovereign bond markets—particularly those of peripheral member states. The latter part of the sample also overlaps with implementation of unconventional monetary policy in the form of largescale bond purchases when these markets became less liquid and harder to hedge. Figure 1 provides a view of the data for the individual sovereign yield changes that was used in the SBBS yield simulation analysis. The negative of yield changes multiplied by 100 is shown. In each case, the conditional volatility and 1% Value-at-Risk are also provided (where the volatility is estimated using a GJR-GARCH(1,1) and the latter is implied by the conditional volatility under normality). It is important to note that the scale of the y -axis is not constant across the panels of this figure.

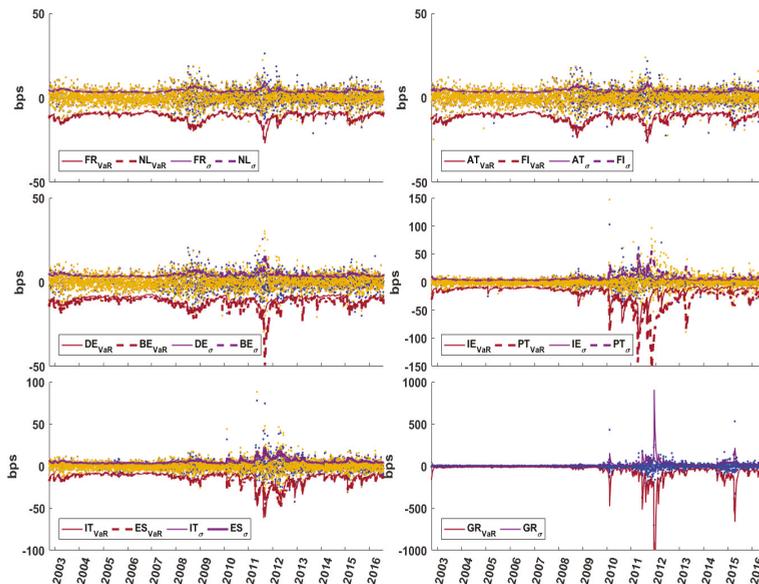


Figure 1. Daily Yield-Change Data 11 Sovereigns (bps). This panel of figures displays the daily data of the individual sovereigns that was used in the simulation analysis that generated the Sovereign Bond-Backed Security yields. Each panel contains data for two sovereigns. The negative of the daily yield changes multiplied by 100 are displayed as dots. In each case, the conditional volatility and 1% Value-at-Risk (VaR) are provided (the conditional volatility is estimated using a GJR-GARCH(1,1) and the VaR is the Value-at-Risk implied using the conditional volatility combined with an assumption of normality). It is important to note that the scale of the y-axis is not constant across the panels of this figure.

More precisely, the SBBS yield estimation method relies on a simulated default-triggering mechanism interacting with an observable market-based proxy for default probability in the underlying securities (in this case, the default probability proxy is simply the *yield premium* in excess of the lowest yield among the sovereigns). The triggering device generates uniformly-distributed triggers on the unit-interval (where all trigger combinations have cross-correlations of 0.6). Whenever these simulated unit-interval triggers exceed the non-default/survival probability, (1 minus the *yield premium*), losses are calculated as though defaults have occurred.

Each simulation produces simulated default losses among the underlying bonds. These are summed and allocated sequentially to the SBBS securities according to their level of subordination (only spilling over to a more senior tranche if simulated losses have exceeded the total par value of all subordinates). The sum of the yield premiums of the national bonds, for each simulated day, is then allocated to the yield premiums of SBBS according to their proportional allocation of simulated default losses. Hence, the likelihood of triggering a simulated default is determined by the size of yield premiums and by how correlated the triggers are. However, risk aversion also has some role in determining the premium.⁹

⁹ The implied risk premium (i.e., yield above the risk-free rate) reflects the risk aversion of the representative market investor on any given day and, hence, may exceed the expected loss anticipated by a risk-neutral investor. This degree of risk aversion enters the simulation and is consequently also reflected consistently in the resulting estimated yields of senior, mezzanine and junior SBBS.

In this way, the probable daily yields on the SBBS components are generated for two different securitisation structures over roughly a 17-year historical period without the need for a structural modelling of the complex dependencies among the underlying sovereigns (e.g., as in Lucas et al. (2017)). This then enables estimation of optimal hedge ratios, hedge effectiveness measures and assessments of the diversification benefits. For reasons of data availability, the simulation is based on yield data for two-, five- and ten-year government bonds of Austria, Belgium, Germany, Spain, Finland, France, Greece, Ireland, Italy, the Netherlands and Portugal, following a weighting scheme based on GDP (averaged over 2006–2015). This basket covers approximately 97.5% of the SBBS volume. As a robustness check, the SBBS yield estimations are re-done using a t-copula instead of the Gaussian copula.

Panels A and B of Figure 2, respectively, depict the time series behaviour of yields on SBBS securities (the derivation of which is discussed above) under two alternative tranching assumptions (70:30 and 70:20:10) while panel C shows yields of a selection of individual sovereigns. The period of the European sovereign debt crisis is highlighted and extends from November 2009, when the Greek government indicated its 2009 deficit projection was being revised upward from 5% to 12.7%, until just after Mario Draghi’s speech making a commitment to ‘do whatever it takes’ to prevent the break-up of the euro in late July 2012. All of the 10-year yield data used in the hedge selection and assessment analysis discussed in the results section has been converted to price and then daily holding period returns assuming a duration of nine years.

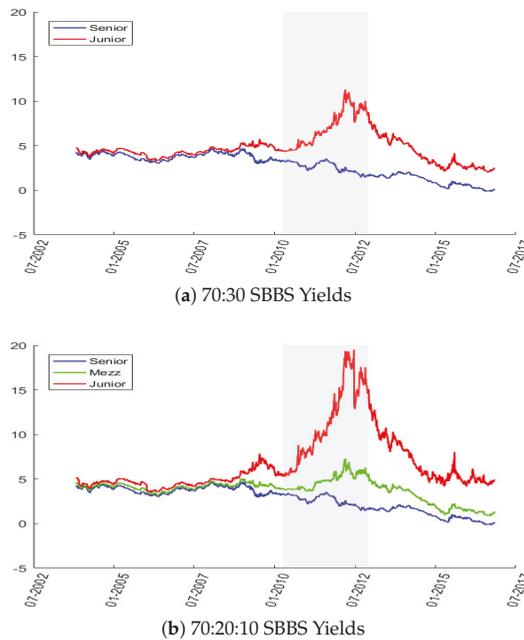


Figure 2. Cont.

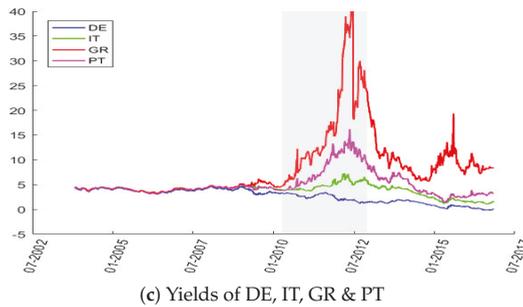


Figure 2. Estimated yields on SBBS tranches & selected sovereigns (%). Note: This panel of figures indicates how SBBS yields are likely to have changed over the sample period. These can be compared with a selection of sovereign yields. The shaded area is the euro area Sovereign Debt Crisis period (November 2009–August 2012).

4.2. Methodology for Optimal Hedge Selection

There is a well developed literature dealing with the selection of optimal hedge ratios. [Chen et al. \(2003\)](#) and [Lien and Tse \(2011\)](#) provide extensive reviews of different theoretical approaches to determining optimal hedge ratios. A prominent approach to optimising a hedge position is based on minimising the variance of the returns on the hedged portfolio without regard to optimising the expected returns in some way. This is a particularly suitable approach in the case of hedging to facilitate market making in sovereigns since the objective is to minimise risk for the reward of earning the bid–ask spread (less costs) rather than to improve returns from the underlying investment.

In the single hedge case, the optimal hedge ratio (see [Ederington 1979](#) and [Baillie and Myers 1991](#)) is simply the negative of the slope coefficient in a regression of the asset return on the hedge instrument return. Composite hedging has sometimes been found to be more effective than relying on a single hedge instrument. This has been claimed in the extant literature for the case of hedging positions in corporate bonds using a combination of the relevant sovereign bond and a futures position in the relevant equity (e.g., [Marcus and Ors 1996](#)), using bonds at particular maturities with futures on a variety of other maturities (e.g., [Morgan 2008](#)) and hedging mortgage backed securities with Treasuries at 2-, 5- and 10-year tenors (e.g., [Koutmos and Pericli 2000](#)). [Garbade \(1999\)](#) also provides an interesting application of a two-asset hedge for a bond. This is similar to the case of hedging with both the Senior and Mezzanine (or Junior) SBBS.

4.3. Measuring Out-Of-Sample Hedge Effectiveness

The hedge selection and measurement carried out in this paper follows [Bessler et al. \(2016\)](#) who conduct a similar analysis in a European sovereign bond context. In that analysis, hedge ratios and effectiveness were estimated using rolling OLS, constant conditional correlation (CCC), dynamic conditional correlation (DCC-GARCH) and a Bayesian based mixture of models. In the current analysis, hedge ratios and hedge effectiveness (using single or multiple SBBS as hedges) are estimated for each of the 11 individual sovereign bonds. Hedge effectiveness in each case is measured by the percentage change in risk achieved through hedging. This is done using two different risk metrics for the hedged and unhedged positions. The first metric is the rolling standard deviations of returns. The second metric is based on Values-at-Risk bounds for the hedged and unhedged positions (i.e., the percentage change in the range between the 5% and 95% Value-at-Risks for hedged and unhedged cases).

Hedge effectiveness is therefore measured by the size of the reduction in risk exposure achieved by hedging (essentially, the risk of the hedged position relative to that of the unhedged position). The optimal hedge ratio and the hedge effectiveness measure typically change through time so hedge effectiveness is assessed on a rolling basis across various sub-samples. [Engle \(2009\)](#) applies a pre-crisis

model of time varying covariance to the problem of hedge selection during the Great Financial Crisis and shows that this improves on static approaches that were the industry norm at that time. The rolling linear regression results discussed below are therefore unlikely to overstate the achievable hedge effectiveness.

To keep the exposition tractable, the results obtained using the rolling linear regressions represent a base case (and arguably a lower bound on effectiveness) and are the main focus of the results presented below.¹⁰ The optimal hedge for each of the 11 sovereigns (with 10 years to maturity) was estimated on a rolling basis for the case of the following seven hedge instruments/combinations: (Senior, Mezzanine, Junior, (Senior and Mezzanine), (Senior and Junior), (Mezzanine and Junior), (Senior, Mezzanine and Junior)). In line with previous literature, the optimal hedge is derived by applying the chosen hedge selection method (e.g., linear regression) over a prior 250 day window and rolling the estimation window forward at regular intervals. The hedge ratio is therefore used in an out-of-sample context for the entire interval between the estimation sub-samples. To keep the exposition tractable, hedge ratio results are presented based on a rolling regression at intervals of 25 days.

5. Results

5.1. Effectiveness of Hedging without Diversification

The hedge effectiveness results, in terms of the percentage change in risk (usually reduction in risk), are shown for each of three sub-sample periods in Tables 1–3 and the daily returns on the hedged and unhedged positions, for the case of hedging with just the senior and for the case of hedging with a mixture of the senior and mezzanine, are shown in Figures 3–5.¹¹

The visual evidence indicates that hedging is generally effective in the pre-Sovereign Debt Crisis period in reducing the volatility of returns (with some isolated exceptions). Hedging is not effective for high-risk sovereigns during the height of the SDC but effectiveness returns to some extent during the recovery. In general, the combined hedge works better than the single hedge in the crises and recovery periods. Figure 3 reveals that hedging is quite consistently effective for three countries (Germany (DE), France (FR) and the Netherlands (NL)). In these cases, the composite hedge seems to eliminate the occasional blips present in the single hedge case. Similarly consistent levels of effectiveness are found for AT and FI (not displayed).

Figure 4 shows the cases of (BE, ES and IT). This clearly reveals how idiosyncratic the experiences of each of the high-risk periphery countries was during the crisis (making them difficult to hedge using SBBS). It is interesting that the composite hedge (senior and mezzanine) works better than the single hedge during the crisis and recovery (apart from one particular day). This tends to improve further with the inclusion of the junior SBBS as a hedge instrument (this more general case is not displayed in the figure but can be seen from the tabulated results discussed below).

Figure 5 shows the more volatile cases of (GR, IE and PT). Here again, there is evidence of hedge ineffectiveness during the crisis with improvement only obvious during the recovery for IE and PT. Again, the composite hedge is better than the single hedge during the recovery for these countries and is particularly good in protecting from the more extreme movements. Although hedging is least effective in reducing risk in the case of the smaller markets, their idiosyncratic risk is more amenable to control using diversification strategies (this is addressed below).

¹⁰ A similar analysis employing dynamic conditional correlation methods, compiling hedge ratios using conditional variances and covariances in line with the relations presented in Appendix A, did not in fact lead to significantly different hedge effectiveness ratios. In addition, the combined use of the methods of Gibson et al. (2017) and a stochastic volatility modelling approach designed to address the presence of isolated outliers, due to Chan and Grant (2016), also failed to change the conclusions drawn from the more straightforward application of a rolling linear regression approach. These alternative methods may however lead to improvements when used at higher frequency with regular updating. Extending the analysis to such a frequency is beyond the scope or needs of the present study.

¹¹ The cases of Austria (AT) and Finland (FI) are not shown above but are quite similar to the case of NL.

Table 1. Hedge effectiveness: Previous to the Sovereign Debt Crisis.

Hedge =	1 SBBS			2-SBBS			3-SBBS
	Snr	Mezz	Jnr	Snr-Mezz	Snr-Jnr	Mezz-Jnr	Snr-Mezz-Jnr
AT(i)	-62	-61	-35	-67	-70	-50	-72
AT(ii)	-73	-72	-35	-77	-82	-57	-84
BE(i)	-65	-63	-36	-72	-75	-52	-77
BE(ii)	-71	-70	-37	-76	-80	-58	-83
DE(i)	-79	-78	-32	-84	-84	-46	-87
DE(ii)	-85	-81	-31	-86	-88	-49	-89
ES(i)	-55	-55	-36	-62	-66	-53	-69
ES(ii)	-62	-61	-36	-69	-73	-58	-75
FI(i)	-70	-69	-35	-72	-75	-46	-76
FI(ii)	-79	-77	-36	-81	-84	-53	-84
FR(i)	-72	-71	-37	-78	-80	-53	-83
FR(ii)	-76	-75	-37	-81	-84	-59	-88
GR(i)	-36	-33	-27	-46	-51	-49	-55
GR(ii)	-46	-44	-33	-60	-60	-58	-67
IE(i)	-42	-40	-26	-47	-51	-39	-52
IE(ii)	-66	-62	-33	-70	-72	-52	-72
IT(i)	-50	-47	-35	-63	-65	-59	-72
IT(ii)	-56	-50	-37	-69	-70	-64	-77
NL(i)	-69	-68	-37	-75	-78	-54	-81
NL(ii)	-77	-75	-36	-80	-83	-58	-86
PT(i)	-50	-48	-34	-59	-63	-54	-67
PT(ii)	-62	-60	-38	-69	-73	-61	-77

Note: Using two different metrics, this table shows the percentage change in risk exposures achieved by hedging a position in a particular sovereign bond using a position in one or more tranches of a 70:20:10 Sovereign Bond Backed Securitisation in the pre-SDC period (this is used as an indicator of *hedge effectiveness* in the text). Issuers of the bonds are indicated at the beginning of each row as follows; AT (Austria), BE (Belgium), DE (Germany), ES (Spain), FI (Finland), FR (France), GR (Greece), IE (Ireland), IT (Italy), NL (Netherlands) and PT (Portugal). The first row for each case contains the percentage change in risk due to hedging where risk is measured as *standard deviation*. The second row for each case contains the percentage change in risk achieved through hedging where risk is measured as the range between the 5th and 95th quantiles of the returns distribution (implying VaR bounds). Columns of results are arranged in three broad categories. The first category concerns the use of a single SBBS tranche in the hedge. The second category concerns the use of a pair of SBBS in hedging. The last category involves the use of all three SBBS in the hedge. The most effective hedge instrument/combination within the first two categories is highlighted with colour.

Table 2. Hedge effectiveness Sov-debt-crisis.

Hedge =	1 SBBS			2-SBBS			3-SBBS
	Snr	Mezz	Jnr	Snr-Mezz	Snr-Jnr	Mezz-Jnr	Snr-Mezz-Jnr
AT(i)	-24	-11	0	-32	-16	4	-26
AT(ii)	-32	-19	-2	-41	-39	-5	-41
BE(i)	-3	-4	-2	-27	10	-16	-20
BE(ii)	-2	-2	0	-27	-10	-17	-29
DE(i)	-68	0	7	-72	-67	4	-71
DE(ii)	-69	4	5	-73	-69	-5	-73
ES(i)	1	10	1	-33	10	-31	-28
ES(ii)	-3	15	5	-29	-13	-34	-35
FI(i)	-52	-7	3	-52	-49	6	-47
FI(ii)	-54	-4	2	-54	-54	4	-55
FR(i)	-23	-12	0	-35	-15	0	-31
FR(ii)	-30	-12	2	-38	-32	2	-38
GR(i)	0	1	0	0	-15	-15	-17
GR(ii)	-4	13	11	2	26	28	23
IE(i)	2	7	2	-3	1	-2	1
IE(ii)	-1	6	3	-5	-8	-7	-6

Table 2. Cont.

Hedge =	1 SBBS			2-SBBS			3-SBBS
	Snr	Mezz	Jnr	Snr-Mezz	Snr-Jnr	Mezz-Jnr	Snr-Mezz-Jnr
IT(i)	0	10	1	-44	18	-39	-37
IT(ii)	2	13	3	-40	-9	-43	-44
NL(i)	-49	-9	2	-48	-46	7	-43
NL(ii)	-53	-6	5	-52	-52	3	-51
PT(i)	1	5	1	-1	1	-2	0
PT(ii)	1	2	1	-5	-10	-8	-9

Note: Using two different metrics, this table shows the percentage change in risk exposures achieved by hedging a position in a particular sovereign bond using a position in one or more tranches of a 70:20:10 Sovereign Bond Backed Securitisation during the SDC period (this is used as an indicator of *hedge effectiveness* in the text). Issuers of the bonds are indicated at the beginning of each row as follows; AT (Austria), BE (Belgium), DE (Germany), ES (Spain), FI (Finland), FR (France), GR (Greece), IE (Ireland), IT (Italy), NL (Netherlands) and PT (Portugal). The first row for each case contains the percentage change in risk due to hedging where risk is measured as *standard deviation*. The second row for each case contains the percentage change in risk achieved through hedging where risk is measured as the range between the 5th and 95th quantiles of the returns distribution (implying VaR bounds). Columns of results are arranged in three broad categories. The first category concerns the use of a single SBBS tranche in the hedge. The second category concerns the use of a pair of SBBS in hedging. The last category involves the use of all three SBBS in the hedge. The most effective hedge instrument/combination within the first two categories is highlighted with colour.

Table 3. Hedge effectiveness Post-Sov-debt-crisis.

Hedge =	1 SBBS			2-SBBS			3-SBBS
	Snr	Mezz	Jnr	Snr-Mezz	Snr-Jnr	Mezz-Jnr	Snr-Mezz-Jnr
AT(i)	-45	-22	0	-47	-49	-10	-49
AT(ii)	-51	-25	0	-53	-57	-14	-56
BE(i)	-44	-26	-2	-48	-53	-13	-52
BE(ii)	-50	-28	-3	-53	-57	-15	-57
DE(i)	-73	-13	4	-74	-73	-8	-75
DE(ii)	-72	-10	4	-73	-73	-7	-74
ES(i)	-2	2	-3	-32	-26	-42	-43
ES(ii)	-4	-6	-4	-29	-28	-41	-43
FI(i)	-52	-16	1	-53	-55	-9	-55
FI(ii)	-59	-18	1	-60	-62	-11	-62
FR(i)	-50	-27	-2	-55	-58	-15	-59
FR(ii)	-54	-28	-2	-56	-61	-16	-61
GR(i)	0	7	7	-8	-8	2	-8
GR(ii)	5	6	8	3	11	17	12
IE(i)	-10	-11	-3	-21	-22	-19	-27
IE(ii)	-14	-17	-5	-29	-28	-23	-35
IT(i)	-3	1	-4	-41	-28	-50	-53
IT(ii)	-7	-5	-4	-41	-34	-52	-54
NL(i)	-53	-18	1	-54	-56	-9	-56
NL(ii)	-60	-18	0	-61	-64	-11	-65
PT(i)	0	2	0	-13	-15	-21	-21
PT(ii)	-1	2	0	-13	-17	-25	-26

Note: Using two different metrics, this table shows the percentage change in risk exposures achieved by hedging a position in a particular sovereign bond using a position in one or more tranches of a 70:20:10 Sovereign Bond Backed Securitisation in the post-SDC period (this is used as an indicator of *hedge effectiveness* in the text). Issuers of the bonds are indicated at the beginning of each row as follows; AT (Austria), BE (Belgium), DE (Germany), ES (Spain), FI (Finland), FR (France), GR (Greece), IE (Ireland), IT (Italy), NL (Netherlands) and PT (Portugal). The first row for each case contains the percentage change in risk due to hedging where risk is measured as *standard deviation*. The second row for each case contains the percentage change in risk achieved through hedging where risk is measured as the range between the 5th and 95th quantiles of the returns distribution (implying VaR bounds). Columns of results are arranged in three broad categories. The first category concerns the use of a single SBBS tranche in the hedge. The second category concerns the use of a pair of SBBS in hedging. The last category involves the use of all three SBBS in the hedge. The most effective hedge instrument/combination within the first two categories is highlighted with colour.

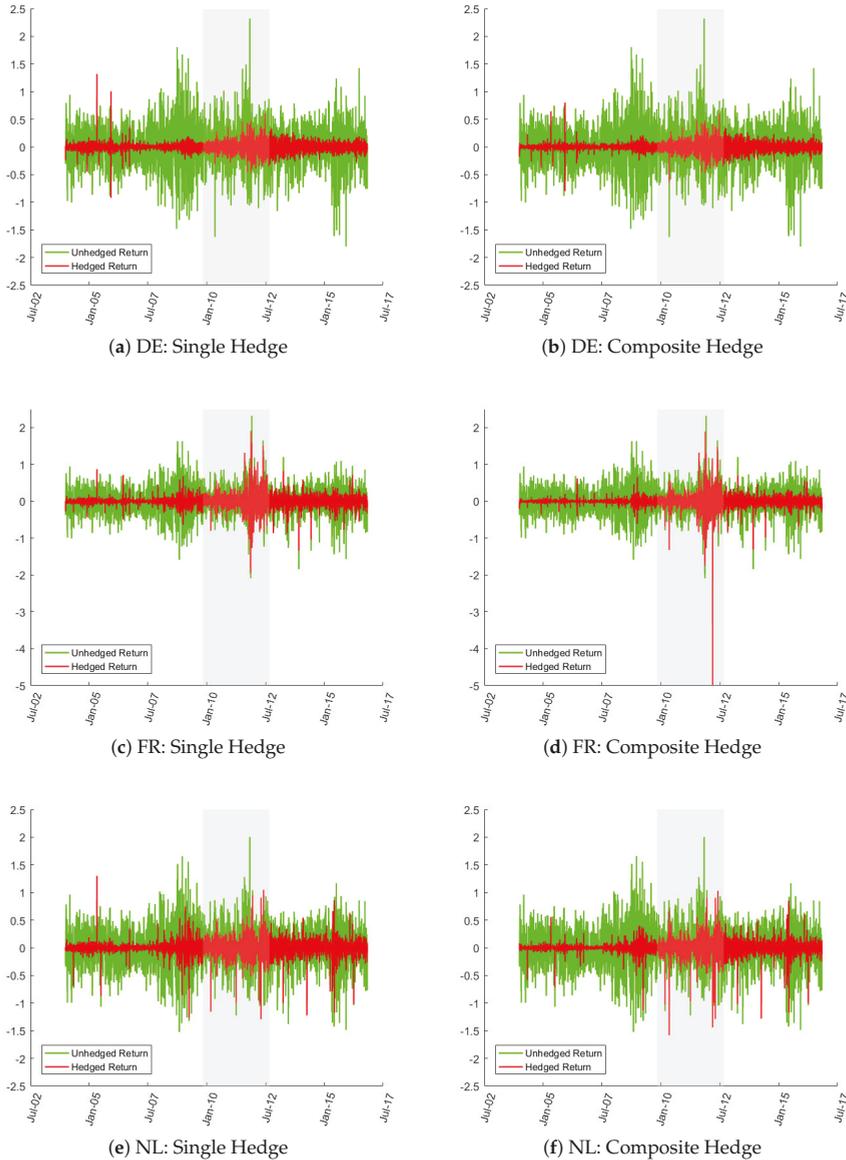


Figure 3. Single and Composite Hedging (DE, FR, NL). Note: This figure facilitates a comparison of the dispersion of returns on hedged long positions in German, French and Dutch sovereign bonds respectively with the dispersion of returns on the same sovereign bonds without hedging. The first column of figures concern the case of hedging the long positions using only the senior SBBS. The second column of figures concerns the case of hedging the individual bond positions with both the senior and mezzanine SBBS. The returns are measured in basis points (left axis) and the bonds considered are those with a 10 year term-to-maturity. The cases of Austria (AT) and Finland (FI) are not shown above but are quite similar to the case of NL.

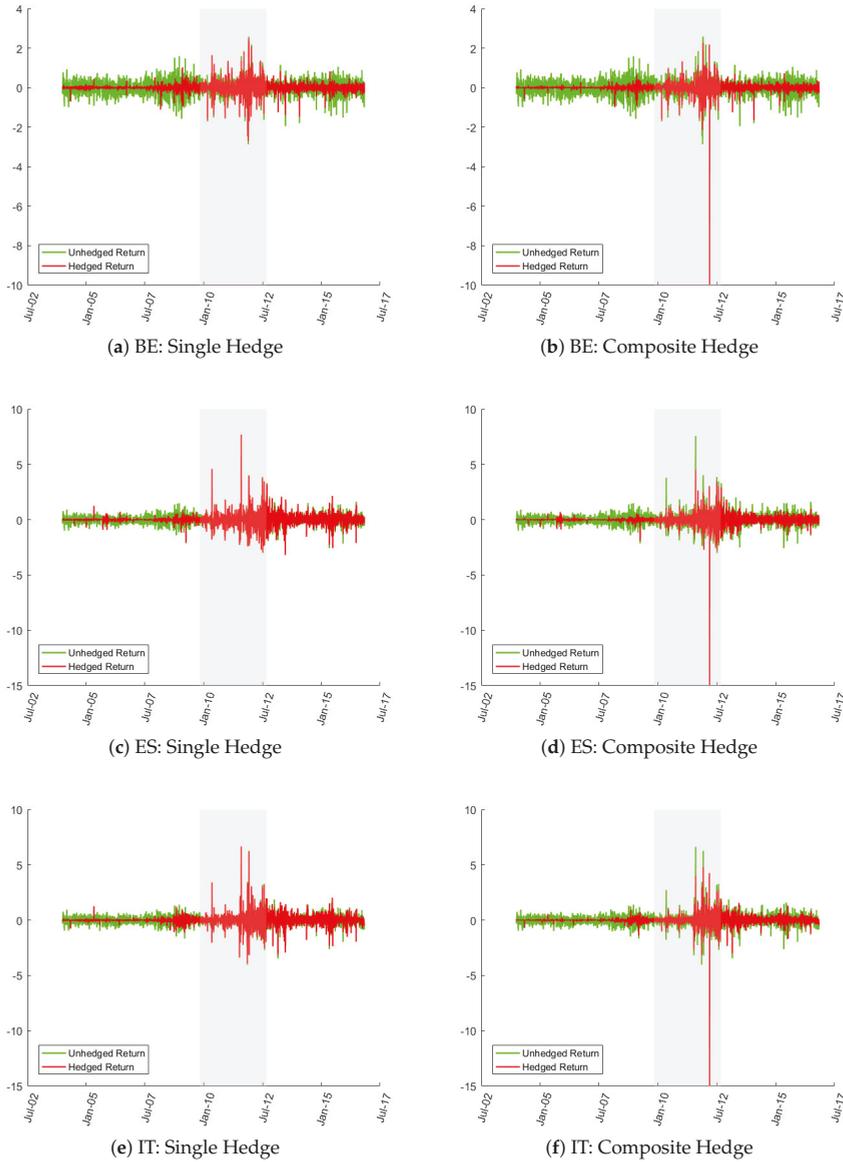


Figure 4. Single and Composite Hedging (BE, ES, IT). Note: This figure facilitates a comparison of the dispersion of returns on hedged long positions in Belgian, Spanish and Italian sovereign bonds respectively with the dispersion of returns on the same sovereign bonds without hedging. The first column of figures concern the case of hedging the long positions using only the senior SBBS. The second column of figures concerns the case of hedging the individual bond positions with both the senior and mezzanine SBBS. The returns are measured in basis points (left axis) and the bonds considered are those with a 10 year term-to-maturity.

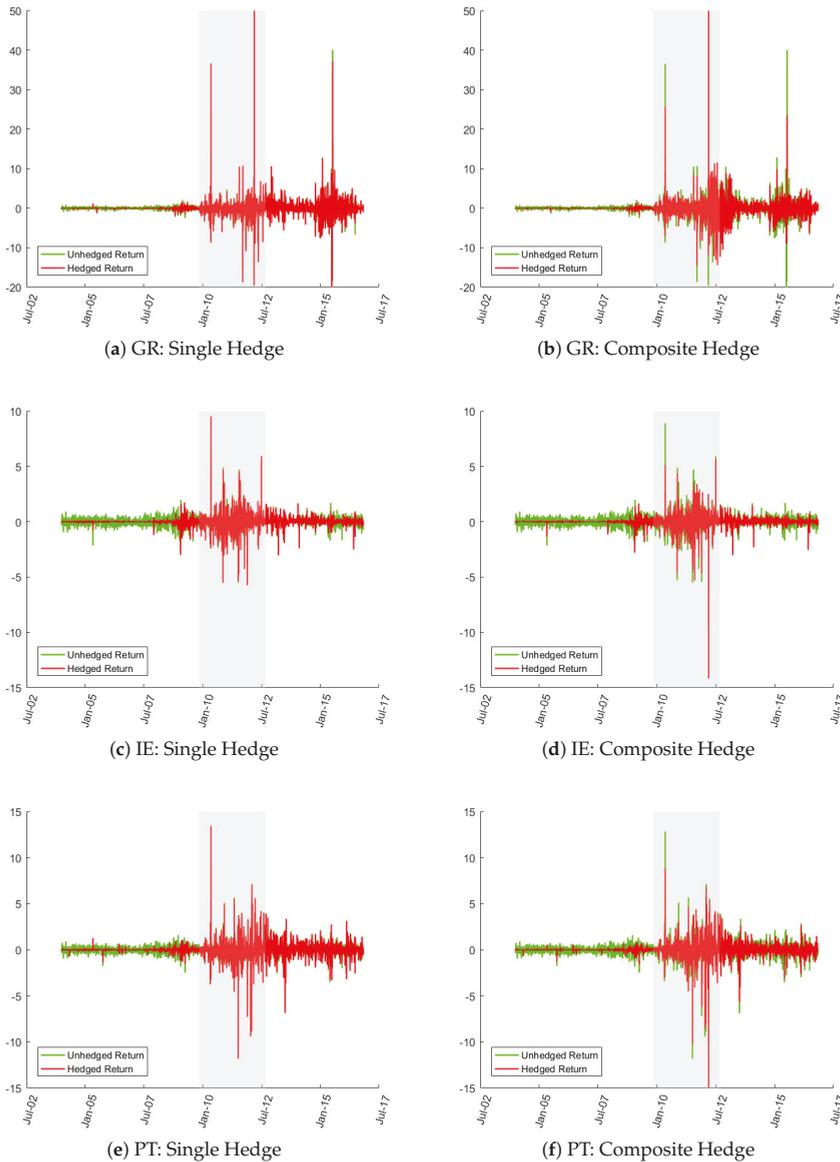


Figure 5. Single and Composite Hedging (GR, IE, PT). Note: This figure facilitates a comparison of the dispersion of returns on hedged long positions in Greek, Irish and Portuguese sovereign bonds respectively with the dispersion of returns on the same sovereign bonds without hedging. The first column of figures concern the case of hedging the long positions using only the senior SBBS. The second column of figures concerns the case of hedging the individual bond positions with both the senior and mezzanine SBBS. The returns are measured in basis points (left axis) and the bonds considered are those with a 10 year term-to-maturity.

The tabulations of risk reductions can provide some more depth to the graphical analysis. The pre-Sovereign Debt Crisis period comparisons are in Table 1 and this is clearly a period where hedge effectiveness is high for all countries (at least for some of the seven hedges). The best hedges for each sovereign are highlighted in colour for the cases of 1-SBBS and 2-SBBS as hedges. In the case of the single hedge, it is the senior-SBBS that gives the best protection (the percentage reduction for the best single hedges—based on standard deviation of daily returns—ranges from 79% for DE to 36% for GR). In almost all cases, the 2-SBBS hedge provides some marginal improvement in hedge effectiveness over the 1-SBBS hedge case (all percentage reductions above 50%). In many cases the 3-SBBS hedge is best overall but this may not always be worthwhile from a cost and operational perspective.

Table 2 shows summary statistics for comparisons of risk changes through hedging during the sovereign debt crisis period. For the 1-SBBS hedge, only DE remains well hedged using the senior SBBS (reducing risk by 68%). Roughly half of the risk is avoided by single-SBBS hedging for the case of FI and NL while the remaining sovereigns are clearly not amenable to single-SBBS hedging in this crisis period. Moving to 2-SBBS or 3-SBBS hedging generally gives rise to some small risk reduction for most sovereigns relative to the single-SBBS case. Table 3 covers the recovery period (from Mario Draghi's London Speech until the end of the last quarter of 2016). By use of composite hedging, it is usually possible to reduce risks by half or more. The exceptions remain GR and PT.

5.2. Post-Hedging Diversification of Risks

Once risks have been hedged, there is potential for dealers to diversify remaining risks by operating simultaneously across many of the sovereign markets. At each moment of a trading day (and at the end of each day), it is likely that a dealer will have a portfolio of outstanding positions in many markets. The individual net hedged-positions are likely to be much smaller than unhedged positions (e.g., if the hedge-ratio is close to 1 then the individual net hedged-positions will be minor). These net positions in individual sovereigns could be subject to further netting (i.e., if some are net-long and others are net-short). Risk reduction from this additional type of netting is not included in the following analysis (i.e., portfolios will be restricted to be hedged long-positions in the components). This most probably implies a significant underestimate of the risk reduction that could be achieved by diversification.

The upper panel of Figure 6 compares returns on a cross-country portfolio of hedged and unhedged positions for the pre-SDC period. In this case, it is assumed that the portfolio results from an equal weighting on the underlying sovereigns. Firstly, it is clear that the portfolio of hedged positions has a much smaller dispersion than the portfolio of unhedged components. Despite the equal weighting of components and the equal capital exposure assumption, diversification reduces risks much more for the portfolio with hedged positions and this is due to the fact that there are mostly idiosyncratic risks surviving in the hedged case while the unhedged case will involve considerable systematic risk.

The lower panel of Figure 6 concerns the more volatile conditions of the SDC and the recovery. In this case, there is not much benefit from the hedging because most risk is actually idiosyncratic. This means that diversification is equally effective in the case of the unhedged portfolio. In general, the risks are much lower than risks for typical single sovereigns during this period. Hedging starts to matter again during the recovery.

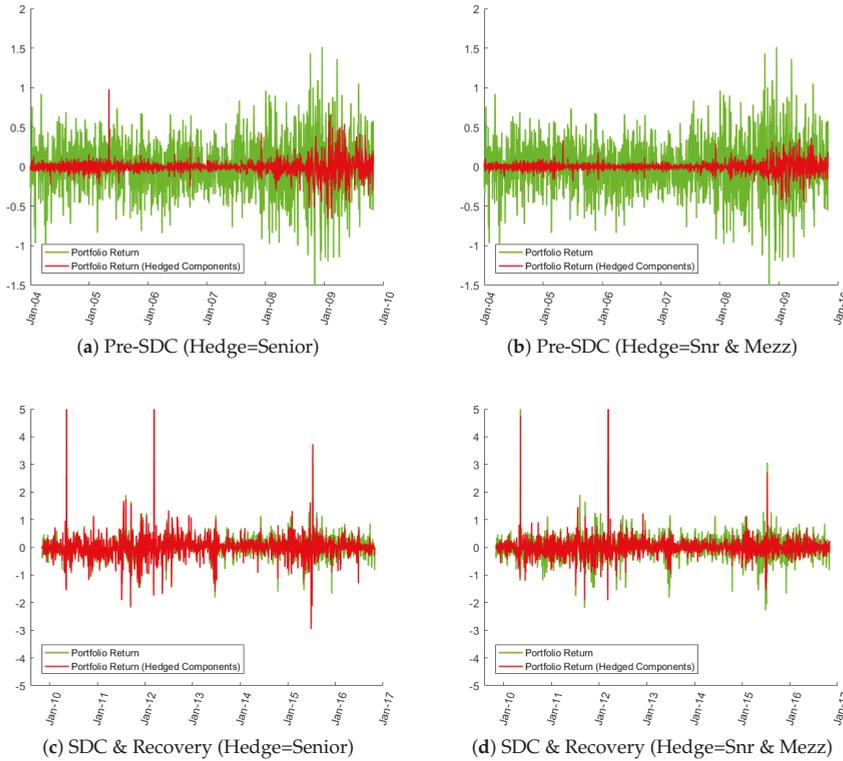


Figure 6. Portfolio returns with and without hedged components. Note: This figure facilitates a comparison of the dispersion of returns on a portfolio of hedged long positions in 11 sovereigns with the dispersion of returns on the same portfolio of long positions without hedging. The first column of figures concern the case of hedging the components of the portfolio of long positions using only the senior SBBS. The second column of figures concerns the case of hedging the individual bond positions with both the senior and mezzanine SBBS. The returns are measured in basis points (left axis) and the bonds considered are those with a 10 year term-to-maturity. The returns are measured in basis points (left axis).

A related comparison is that of the dispersion of returns on the safest single sovereign asset and those of the portfolio of hedged sovereign positions (the German 10 year Bund returns are used for this discussion, but these are almost indistinguishable from the returns on the similarly safe senior SBBS). Figure 7 displays these returns through time for the pre-SDC period and the sample from the start of the SDC through to the more recent recovery. It is very obvious that the portfolio of hedged sovereigns has an extremely low risk level when compared with the most safe sovereign (especially in the pre-SDC period). This demonstrates the effectiveness of diversification in reducing risks when risks are idiosyncratic and not too extreme. The effectiveness of diversification is compromised from the start of the SDC, but there is still some significant reduction in risk for the majority of the sample (with a small number of extreme outliers). For normal risk levels, diversification is a strong driver of risk reduction.

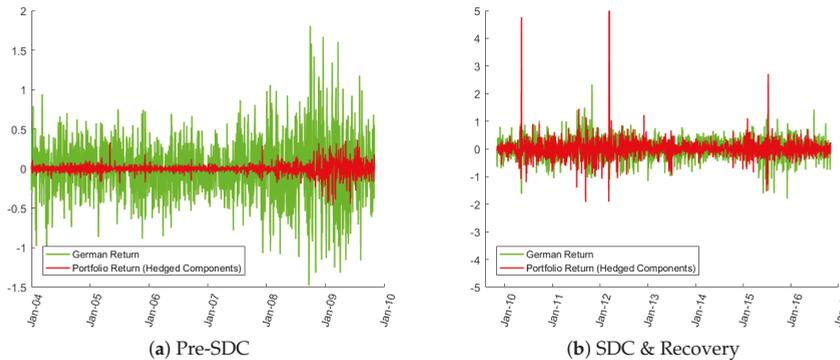


Figure 7. German 10-year returns compared with those from an 11-country portfolio of sovereigns hedged with SBBS. Note: This figure facilitates a comparison of the dispersion of returns on a portfolio of hedged long positions in 11 sovereigns with the dispersion of returns on a single sovereign bond that is widely considered to be the safest sovereign bond investment in the euro area (namely the German Bund). The long positions are hedged using the senior SBBS. The returns are measured in basis points (left axis) and the bonds considered are those with a 10 year term-to-maturity. (a) shows the comparison for the period previous to the Sovereign Debt Crisis while (b) covers the period of the Sovereign Debt Crisis and recovery. The hedged portfolio has returns with a dispersion which is much lower than the safe haven asset in the non-crisis sub-sample. Even in the crisis and recover periods, the hedged portfolio compares favourably with the safe asset investment in terms of dispersion of returns.

Table 4 broadens the analysis of the relative risks following diversification to the case of a size-weighted portfolio and to other maturities. These results pertain to the case where the senior SBBS is used as the hedge instrument.¹² The tabulated results for the 10-year bonds confirm what was clear from the graphical analysis. In the pre-SDC period, most risk measures indicate a reduction of around 70% in risk due to a combination of diversification and hedging relative to diversification on its own. It also confirms that risks after hedging and diversifying are less than a third of those faced by investing in the lowest risk sovereign (close to what is the maximum reduction possible when investing the same capital in 11 assets with similar idiosyncratic risks rather than 1). In this period (for the 10 year term), there is little difference between the equal and size-weighted portfolio results.

For the 10 year term, there is little benefit from hedging before diversifying during the Sovereign Debt Crisis period (the ratio of the risks on portfolio returns when components are hedged is roughly equal to the risks with unhedged components which is tantamount to saying that almost all risks are idiosyncratic). Diversification reduces risks again in the Recovery period and this is particularly true for the size-weighted portfolio. Risk is relatively high for the equally-weighted portfolio compared with the German bond for the Sovereign Debt Crisis period, but otherwise risk is similar to or less than that of the German for the SDC and Recovery, respectively.

In the case of the 5-year term, risks are reduced by around 45% in the pre-SDC period via a combination of hedging and diversification relative to just diversification (the extra risk reduction achieved in the 2 year case is similar to that of the 10 year). As with the 10 year case, hedging does not achieve significant extra benefit in terms of risk reduction for the 5-year bonds during the SDC (and this is also true of the two year term-to-maturity). There is about a ten percentage point difference in the risk reduction due to hedging during the Recovery between the 2 and 5 year terms with more

¹² The weights used in the analysis are related to the size of the individual sovereign float relative to that of the total of 11 sovereigns in the euro area market (weights are provided in the notes for Table 4).

reduction achieved in the 5 year category. As for comparison with the German sovereign, an equally weighted portfolio of sovereigns has risk of 1.52 and 3.52 times the volatility of the German bond (for the 5 and 2 year case, respectively) during the SDC (the associated relative VaR measures and the size-weighted comparisons show a smaller difference in these risks, but, in all cases, risks are greater than holding just the German). The Recovery period does not appear to be as positive for risk reduction through the combination of hedging and diversification for the 2 year term. However, portfolio risks remain low and are not much greater than those of the lowest risk sovereign.

Table 4. Diversification and hedge effectiveness.

Weighting =	Pre-Crisis		Sov Debt Crisis		Recovery	
	Equal	Size	Equal	Size	Equal	Size
10-Year						
EA(Hedged)/EA (i)	-68	-71	0	-3	-12	-24
EA(Hedged)/EA (ii)	-73	-74	-4	-7	-18	-31
EA(Hedged)/DE (i)	-71	-73	8	-1	-4	-24
EA(Hedged)/DE (ii)	-76	-76	-2	-14	-11	-29
5-Year						
EA(Hedged)/EA (i)	-47	-43	1	-1	-17	-15
EA(Hedged)/EA (ii)	-46	-44	0	1	-20	-21
EA(Hedged)/DE (i)	-50	-46	52	25	-21	-3
EA(Hedged)/DE (ii)	-50	-46	34	12	-27	-10
2-Year						
EA(Hedged)/EA (i)	-62	-64	1	-1	-5	-5
EA(Hedged)/EA (ii)	-69	-72	1	4	-7	-10
EA(Hedged)/DE (i)	-70	-67	252	81	21	47
EA(Hedged)/DE (ii)	-76	-74	186	64	15	38

Note: Using two different metrics, this table shows the percentage change in risk exposures achieved by hedging and diversifying (using the optimal hedge position in one or more tranches of a 70:20:10 Sovereign Bond Backed Securitisation). The portfolio of assets being hedged and diversified involves the 10 year bonds issued by; AT (Austria), BE (Belgium), DE (Germany), ES (Spain), FI (Finland), FR (France), GR (Greece), IE (Ireland), IT (Italy), NL (Netherlands) and PT (Portugal). The first row for each case contains the percentage change in risk due to hedging where risk is measured as *standard deviation*. The second row for each case contains the percentage change in risk achieved through hedging where risk is measured as the range between the 5th and 95th quantiles of the returns distribution (implying VaR bounds). The risk reduction is measured relative to the unhedged portfolio and relative to an investment in just the German Bund which is considered the safe haven asset. Columns of results permit a comparison across three sub-sample periods and for risk reductions achieved by diversifying across an equally-weighted and size-weighted portfolio, respectively.

5.3. Summary of Results for Extensions

The possibility that similar hedge (and diversification) strategies might be just as good as SBBS has been ignored. The effectiveness of hedging using futures (in particular, German Bund and Italian BTP bond futures) is addressed in Appendix B. The principal finding here is that the senior SBBS is a much better hedge than Bund futures (even ignoring the fact that hedging with futures involves other additional costs and basis risks). Additionally, where SBBS are ineffective as hedges for smaller sovereigns (such as IE and PT), the futures are also very ineffective so there is no clear superiority in terms of hedge effectiveness using futures rather than SBBS.

Appendix B also contains details of the generalisation of the analysis to other maturities and to the case where SBBS yields are generated using a multivariate t-copula to trigger defaults. The outcome of these extensions is straightforward to summarise. Firstly, extending to shorter maturities leads to a general reduction in hedge effectiveness on average in the pre-crisis period and a less pervasive reduction in hedge effectiveness for the two-year maturity in the recovery period. Ultimately, this does not impact greatly on the risks that remain after diversification. Secondly, since the t-copula produces a higher incidence of extreme losses that more often spillover to mezzanine and senior tranches, the yield movements that result are less correlated with those of the German sovereign bonds than is the case with the Gaussian copula. The senior tranche therefore becomes less effective as a hedge for German bonds. Otherwise, the changes in hedge effectiveness due to use of the t-copula are insignificant.

6. Conclusions

This paper assesses how the existence of sovereign bond-backed securities affects dealer intermediation in individual euro area sovereigns. An arbitrage relation is used to demonstrate that the hedging of inventory risks with SBBS eliminates most systematic inventory holding risks. Furthermore, it is asserted that diversification of intermediation activities across countries, in combination with hedging, then produces further reductions in exposures. These assertions are tested using estimated SBBS yields and the findings are generally positive. Even if one assumes a similar capital exposure under hedging and not-hedging, there is still a marked reduction in risks through hedging and diversification. Overall, assuming regulation does not excessively penalise netting of inventory positions in different sovereign and SBBS, and if SBBS are sufficiently liquid, a significant improvement in trading costs across all European sovereign debt markets seems plausible.

Risk reduction through diversification (after hedging) was assessed using equal-weighted and size-weighted portfolios of long positions in the bonds from the underlying sovereigns. This could be misleading since it is quite likely that actual inventories would be some mixture of long and short positions. This simplification will tend to underestimate the risk-reducing benefits of diversification. Since holding period investment returns in sovereign markets are predominantly positively correlated (except when there is a pronounced flight to safety), a combination of long and short positions will lead to more netting of valuation changes and this will most likely reduce risk relative to what is reported for the present analysis. This represents a fruitful avenue for future investigation.

A further useful extension of the above analysis would be to assess whether there are liquidity spillovers among the SBBS tranches. In fact, these spillovers have already been identified in terms of the incremental hedge effectiveness that was achieved when portfolios that included all three SBBS were analysed. However, in the developmental phase of the SBBS market, liquidity spillovers between the SBBS themselves could be important for liquidity of the junior tranche which will have the smallest issuance volume among the three. It seems likely that the senior SBBS would be helpful to dealers in providing quotes in the mezzanine market (purely via hedging). A spillover from the mezzanine to the junior is then possible. Preliminary analysis of the correlations between the tranches under most circumstances suggests that these hedging avenues for liquidity spillovers would be positive. Of course, in crisis circumstances, flight to safety could disrupt correlations more dramatically than can be anticipated by dealers and this would undermine the reliability of risk reduction strategies based purely on hedging one SBBS with another. This is worthy of further investigation.

It is important to acknowledge that providing liquidity by relying on a parallel market for hedging requires adequate funding liquidity, regulation that permits the netting of inventory positions and broad-based diversification by dealers. Any systemic contraction in availability of funding liquidity is likely to disproportionately affect liquidity in markets that depend on hedging (see [Brunnermeier and Pedersen 2009](#)). It is important therefore that dealers have adequate capital to withstand relatively large shocks.¹³ Regulation that prevents netting of positions of intermediaries would constrain positive liquidity spillover effects. Similarly, large changes in correlations can magnify risks during a crisis so trading systems need to be dynamically flexible and capable of managing such complexity. This may increase operational costs.

One important policy implication of the analysis above is that primary dealers will seek to have a presence in more markets. This follows from the fact that much of the risk at local level cannot be hedged using SBBS but is easily diversified. Dealers with, on average, a more diversified portfolio will face lower risks and will be able to out-price non-diversified dealers. Most markets in Europe have a majority of dealers with diversified market making activities, but some additional diversification is

¹³ It has been shown by [Baranova et al. \(2017\)](#) that recent tightening of capital and leverage requirements of financial intermediaries has damaged liquidity provision during calm markets' conditions, but it has helped to protect liquidity during crises' circumstances. Getting the balance right is therefore crucial.

likely to occur. There is obviously a trade-off here between specialisation, which helps to make price discovery more efficient, and diversification for risk management purposes.

Funding: This research received no external funding.

Acknowledgments: I am grateful to Sam Langfield, Marco Pagano, Richard Portes, Philip Lane, Javier Suárez, Kitty Moloney, Patrick Haran, Angelo Rinaldo, Thomas Reininger, Martin Phul and members of the ESRB High-Level Task Force on Safe Assets for helpful comments. Views expressed are those of the author and do not necessarily represent the views of the task force. Technical assistance from Paul Huxley at Matlab is also gratefully acknowledged.

Conflicts of Interest: The author declares no conflict of interest.

Appendix A. Optimal Hedging

The optimal hedge ratio is the one that minimises the variance of the hedged portfolio's return. To hedge quantity Q_1 of an asset using quantities of representative combinations of hedge instruments denoted Q_2 , Q_3 and Q_4 gives the following hedge ratios (where ρ_{ij} is the correlation between returns on assets i and j , and σ_i is the standard deviation of returns on asset i).

One hedge instrument:
$$\frac{Q_2}{Q_1} = -\frac{\rho_{12}\sigma_{R_1}\sigma_{R_2}}{\sigma_{R_2}^2},$$

Two hedge instruments:
$$\frac{Q_2}{Q_1} = -\frac{(\rho_{12}-\rho_{13}\rho_{23})\sigma_1\sigma_2}{(1-\rho_{23}^2)\sigma_2^2} \quad \text{and} \quad \frac{Q_3}{Q_1} = -\frac{(\rho_{13}-\rho_{12}\rho_{23})\sigma_1\sigma_3}{(1-\rho_{23}^2)\sigma_3^2},$$

Three hedge instruments:

$$\frac{Q_2}{Q_1} = -\frac{(1-\rho_{34}^2)\rho_{12} - (\rho_{23} - \rho_{24}\rho_{34})\rho_{13} + (\rho_{23}\rho_{34} - \rho_{24})\rho_{14}}{(1-\rho_{23}^2 - \rho_{24}^2 - \rho_{34}^2 + 2\rho_{23}\rho_{24}\rho_{34})\sigma_2},$$

$$\frac{Q_3}{Q_1} = -\frac{(1-\rho_{24}^2)\rho_{13} - (\rho_{23} - \rho_{24}\rho_{34})\rho_{12} + (\rho_{23}\rho_{24} - \rho_{34})\rho_{14}}{(1-\rho_{23}^2 - \rho_{24}^2 - \rho_{34}^2 + 2\rho_{23}\rho_{24}\rho_{34})\sigma_3},$$

$$\frac{Q_4}{Q_1} = -\frac{(1-\rho_{23}^2)\rho_{14} - (\rho_{34} - \rho_{24}\rho_{23})\rho_{13} + (\rho_{23}\rho_{34} - \rho_{24})\rho_{12}}{(1-\rho_{23}^2 - \rho_{24}^2 - \rho_{34}^2 + 2\rho_{23}\rho_{24}\rho_{34})\sigma_4}.$$

Correlations used in the the above calculations can be derived from dynamic estimates of variances as described in [Gibson et al. \(2017\)](#). For example, since the variance of $(X + Y) = \text{variance}(X) + \text{Variance}(Y) + 2 \text{Covariance}(X,Y)$ the Covariance can be constructed from a rearrangement of estimates of variance $(X + Y)$, variance (X) and Variance (Y) .

Appendix B. Robustness

Appendix B.1. Comparing with Futures as Hedge

The comparison of hedge effectiveness using SBBS versus BTP and/or Bund futures can be considered by reference to [Table A1](#). The first notable outcome is that the senior SBBS is better than the Bund future even in the case of hedging German 10-year bond risk. The returns on a portfolio of Bund and an optimal hedge position in Bund futures has a standard deviation of around 67% and 64% respectively of the unhedged position in the SDC period and the recovery, respectively. The portfolio involving the senior SBBS by comparison has a relative standard deviation of 32% and 27% in these periods (the relative VaR in both periods is also much lower when the SBBS is used to hedge). Interestingly, the BTP future is a better hedge for the Italian bond than any of the SBBS individually, but a combination of SBBS achieves similar hedge effectiveness.

Overall, the Bund futures contract is a weaker hedge for most sovereign bond positions compared with SBBS. Where SBBS are ineffective as hedges for smaller sovereigns (such as IE and PT), the futures are also very ineffective. These sovereigns have a lot of idiosyncratic unhedgable risk, but their risks are significantly reduced through diversification.

Table A1. Futures-based hedge effectiveness SDC and Recovery.

Hedge =	Sov Debt Crisis			Recovery		
	BUND	BTP	BUND+BTP	BUND	BTP	BUND+BTP
AT(i)	-28	7	-27	-17	-3	-19
AT(ii)	-26	13	-24	-23	-7	-29
BE(i)	-1	-11	-19	-10	-6	-16
BE(ii)	10	2	-12	-16	-5	-22
DE(i)	-33	3	-33	-36	-3	-35
DE(ii)	-32	6	-32	-44	-4	-44
ES(i)	2	-36	-33	-2	-24	-23
ES(ii)	-2	-33	-35	-4	-26	-26
FI(i)	-34	5	-33	-32	-3	-32
FI(ii)	-38	0	-38	-41	-7	-44
FR(i)	-22	6	-21	-16	-4	-19
FR(ii)	-18	12	-22	-20	-8	-27
GR(i)	3	13	17	0	0	0
GR(ii)	3	18	19	3	0	2
IE(i)	0	-5	-6	-1	0	-1
IE(ii)	-4	-2	-5	-4	-1	-3
IT(i)	3	-52	-49	-3	-46	-45
IT(ii)	7	-57	-59	-5	-59	-58
NL(i)	-33	5	-32	-29	-2	-29
NL(ii)	-29	3	-31	-37	-3	-38
PT(i)	0	-4	-5	0	-3	-2
PT(ii)	3	-8	-5	1	-5	-6

Note: Using two different metrics, this table shows the percentage change in risk exposures achieved by hedging a position in a particular sovereign bond using a position in futures on Bunds, BTPs and a combination of both. Issuers of the bonds are indicated at the beginning of each row as follows; AT (Austria), BE (Belgium), DE (Germany), ES (Spain), FI (Finland), FR (France), GR (Greece), IE (Ireland), IT (Italy), NL (Netherlands) and PT (Portugal). The first row for each case contains the percentage change in risk due to hedging where risk is measured as *standard deviation*. The second row for each case contains the percentage change in risk achieved through hedging where risk is measured as the range between the 5th and 95th quantiles of the returns distribution (implying VaR bounds). Columns of results permit a comparison of the risk reductions for the different hedges across two sub-sample periods.

Appendix B.2. Hedge Effectiveness at Other Maturities

Table A2 presents a comparison of hedge effectiveness at the 10-year term-to-maturity with those at five and two years to maturity (for conciseness only the results for hedging with all three SBBS are shown). In the pre-crisis period hedge effectiveness tends to decline as term shortens. The standard deviation of returns on the hedged positions relative to that of the unhedged positions (Rows (i)) increases on average from 0.28 at the 10 year term to 0.36 and 0.49 at the 5- and 2-year terms respectively. The relative value-at-risk comparison (Rows (ii)) increase on average less dramatically than the relative standard deviations (from 0.20 to 0.25 and 0.36, respectively). The average increase in these ratios in the pre-crisis period is a reasonably good indication of what happens at the individual sovereign level (the case of Finland is somewhat of an outlier).

The middle segment of Table A2 depicts the hedge effectiveness measures across terms-to-maturity for the period of the Sovereign Debt Crisis. Here, it is not so clear that any significant change occurs across the different terms on average, but there is more heterogeneity across sovereigns. The average of the ratio of standard deviations of returns on the hedged position relative to the unhedged, shown in the second last row of the table, move from 0.71 to 0.67 and then to 0.72. Relative VaRs also remain quite flat, moving from 0.67 at the 10-year term to 0.68 and 0.73 at the 5- and 2-year terms, respectively. The declines in these ratios tend to be more acute for ES, IT and PT. Significant increases occur (particularly for the 2-year term) for BE, DE, FI and NL. The hedge effectiveness results for the Recovery period show similar levels on average of the two main risk ratios between the 5- and 10-year maturities. A slightly sharper rise occurs in these ratios for the 2-year term and this is mainly explained by the FR case and two of the smaller sovereign markets (BE and FI).

Table A2. Hedge effectiveness: 10-, 5- & 2-year terms-to-maturity

Term =	Pre-Crisis			Sov Debt Crisis			Recovery		
	10 Year	5 Year	2 Year	10 Year	5 Year	2 Year	10 Year	5 Year	2 Year
AT(i)	-72	-65	-47	-26	-29	-23	-49	-44	-24
AT(ii)	-84	-80	-58	-41	-40	-30	-56	-51	-24
BE(i)	-77	-67	-56	-20	-23	-15	-52	-35	-15
BE(ii)	-83	-81	-76	-29	-26	-20	-57	-45	-10
DE(i)	-87	-85	-75	-71	-69	-48	-75	-72	-68
DE(ii)	-89	-88	-82	-73	-69	-55	-74	-73	-67
ES(i)	-69	-67	-62	-28	-44	-42	-43	-50	-43
ES(ii)	-75	-80	-77	-35	-38	-36	-43	-48	-37
FI(i)	-76	-55	-26	-47	-45	-20	-55	-41	-27
FI(ii)	-84	-80	-30	-55	-56	-26	-62	-56	-24
FR(i)	-83	-81	-75	-31	-33	-29	-59	-49	-31
FR(ii)	-88	-86	-82	-38	-38	-35	-61	-52	-27
GR(i)	-55	-45	-34	-17	0	1	-8		
GR(ii)	-67	-60	-61	23	31	42	12		
IE(i)	-52	-40	-17	1	-13		-27	-14	
IE(ii)	-72	-40	-27	-6	-10		-35	-19	
IT(i)	-72	-64	-59	-37	-59	-62	-53	-54	-60
IT(ii)	-77	-70	-73	-44	-51	-61	-54	-55	-57
NL(i)	-81	-71	-54	-43	-42	-25	-56	-47	-39
NL(ii)	-86	-84	-72	-51	-53	-42	-65	-57	-36
PT(i)	-67	-59	-58	0	-9	-17	-21	-13	-8
PT(ii)	-77	-76	-69	-9	-5	-3	-26	-11	1
Avg(i)	-72	-64	-51	-29	-33	-28	-45	-42	-35
Avg(ii)	-80	-75	-64	-33	-32	-27	-47	-47	-31

Note: Using two different metrics, this table shows the percentage change in risk exposures achieved by hedging a position in a particular sovereign bond using a position in one or more tranches of a 70:20:10 Sovereign Bond Backed Securitisation (three maturities are considered; 2-year, 5-year and 10-year). Issuers of the bonds are indicated at the beginning of each row as follows; AT (Austria), BE (Belgium), DE (Germany), ES (Spain), FI (Finland), FR (France), GR (Greece), IE (Ireland), IT (Italy), NL (Netherlands) and PT (Portugal). The first row for each case contains the percentage change in risk due to hedging where risk is measured as *standard deviation*. The second row for each case contains the percentage change in risk achieved through hedging where risk is measured as the range between the 5th and 95th quantiles of the returns distribution (implying VaR bounds). Columns of results permit a comparison of the risk reduction for three bond maturities and across three sub-sample periods.

Appendix B.3. Hedge Effectiveness under Higher Incidence of Extreme Losses

Table A3 compares hedge effectiveness using a t-copula rather than a Gaussian copula to trigger defaults. There is almost no difference for any sovereign (or on average) in the hedge effectiveness measures over the pre-crisis sample. The largest increase in the ratio of risks occurs for the case of GR (showing a very minor 4-point rise). For the Sovereign Debt Crisis and Recovery periods, a similar small change is apparent for all sovereign bonds except those of Germany (and Finland for the SDC period).

Table A3. Hedge effectiveness: Gaussian versus T-copula default trigger.

	Pre-Crisis		Sov Debt Crisis		Recovery	
	Gaussian	T-Dist	Gaussian	T-Dist	Gaussian	T-Dist
AT(i)	-72	-72	-26	-24	-49	-46
AT(ii)	-84	-83	-41	-38	-56	-52
BE(i)	-77	-77	-20	-19	-52	-49
BE(ii)	-83	-82	-29	-26	-57	-54
DE(i)	-87	-81	-71	-45	-75	-56
DE(ii)	-89	-83	-73	-48	-74	-53
ES(i)	-69	-68	-28	-34	-43	-40
ES(ii)	-75	-76	-35	-38	-43	-40
FI(i)	-76	-75	-47	-38	-55	-48
FI(ii)	-84	-83	-55	-44	-62	-55
FR(i)	-83	-84	-31	-26	-59	-57
FR(ii)	-88	-88	-38	-44	-61	-59
GR(i)	-55	-51	-17	-17	-8	0
GR(ii)	-67	-63	23	29	12	27
IE(i)	-52	-50	1	0	-27	-26
IE(ii)	-72	-73	-6	-9	-35	-34
IT(i)	-72	-70	-37	-40	-53	-51
IT(ii)	-77	-71	-44	-40	-54	-55
NL(i)	-81	-81	-43	-37	-56	-51
NL(ii)	-86	-87	-51	-46	-65	-57
PT(i)	-67	-65	0	2	-21	-18
PT(ii)	-77	-76	-9	-14	-26	-21
Avg(i)	-72	-70	-29	-25	-45	-40
Avg(ii)	-80	-79	-33	-29	-47	-41

Note: Using two different metrics, this table shows the percentage change in risk exposures achieved by hedging a position in a particular sovereign bond using a position in one or more tranches of a 70:20:10 Sovereign Bond Backed Securitisation where the tranche yields have been estimated with a Gaussian and a t-Copula simulator. Issuers of the bonds are indicated at the beginning of each row as follows: AT (Austria), BE (Belgium), DE (Germany), ES (Spain), FI (Finland), FR (France), GR (Greece), IE (Ireland), IT (Italy), NL (Netherlands) and PT (Portugal). The first row for each case contains the percentage change in risk due to hedging where risk is measured as *standard deviation*. The second row for each case contains the percentage change in risk achieved through hedging where risk is measured as the range between the 5th and 95th quantiles of the returns distribution (implying VaR bounds). Columns of results permit a comparison of the risk reduction using two ways of estimating the tranche yields and across three sub-sample periods.

References

Amihud, Yakov, and Haim Mendelson. 1980. Dealership Market: Market Making with Inventory. *Journal of Financial Economics* 8: 31–53. [CrossRef]

Baillie, Richard T., and Robert J. Myers. 1991. Bivariate GARCH estimation of the optimal commodity futures hedge. *Journal of Applied Econometrics* 6: 109–24. [CrossRef]

Basel Committee on Banking Supervision. 2017. *Regulatory Treatment of Sovereign Exposures*. Basel: Basel Committee on Banking Supervision.

Bessler, Wolfgang, Alexander Leonhardt, and Dominik Wolff. 2016. Analyzing Hedging Strategies for Fixed Income Portfolios: A Bayesian Approach for Model Selection. *International Review of Financial Analysis*. Available online: <http://dx.doi.org/10.2139/ssrn.2429344> (accessed on 8 April 2019).

Brunnermeier, Markus K., and Lasse Heje Pedersen. 2009. Market Liquidity and Funding Liquidity. *The Review of Financial Studies* 22: 2201–38. [CrossRef]

Brunnermeier, Markus K., Sam Langfield, Marco Pagano, Ricardo Reis, Stijn Van Nieuwerburgh, and Dimitri Vayanos. 2016. The sovereign-bank diabolic loop and ESBies. *American Economic Review Papers and Proceedings* 106: 508–12. [CrossRef]

Brunnermeier, Markus K., Sam Langfield, Marco Pagano, Ricardo Reis, Stijn Van Nieuwerburgh, and Dimitri Vayanos. 2017. ESBies: Safety in the Tranches. *Economic Policy* 32: 175–219. [CrossRef]

Chan, Joshua CC, and Angelia L. Grant. 2016. Modeling energy price dynamics: GARCH versus stochastic volatility. *Energy Economics* 54: 182–89. [CrossRef]

- Chen, Sheng-Syan, Cheng-few Lee, and Keshab Shrestha. 2003. Futures Hedge Ratios: A Review. *Quarterly Review of Economics and Finance*. 43: 433–65. [[CrossRef](#)]
- Copeland, Thomas E., and Dan Galai. 1983. Information Effects on the Bid-Ask Spread. *Journal of Finance* 38: 1457–69. [[CrossRef](#)]
- Demsetz, Harold. 1968. The Cost of Transacting. *Quarterly Journal of Economics* 82: 33–53. [[CrossRef](#)]
- De Sola Perea, M., Peter G. Dunne, Martin Puhl, and Thomas Reiningger. 2017. Sovereign Bond Backed Securities: A VAR-for-VaR & Marginal Expected Shortfall Assessment. ESRB Working Paper Series, European Systemic Risk Board (ESRB), Frankfurt, Germany.
- Dunne, Peter G., Michael J. Moore, and Richard Portes. 2007. Benchmark Status in Fixed-Income Asset Markets. *Journal of Business, Finance & Accounting* 34: 1615–34. [[CrossRef](#)]
- Easley, David, and Maureen O'hara. 1987. Price, Trade Size, and Information in Securities Markets. *Journal of Financial Economics* 19: 69–90. [[CrossRef](#)]
- Ederington, Louis. 1979. The Hedging Performance of the New Futures Markets. *Journal of Finance* 34: 157–70. [[CrossRef](#)]
- Engle, Robert. 2009. *Anticipating Correlations*. Princeton: Princeton University Press.
- ESRB High-Level Task Force on Safe Assets. 2018. Sovereign bond-backed securities: A feasibility study. Frankfurt: European Systemic Risk Board.
- Farhi, Emmanuel, and Jean Tirole. 2018. Deadly embrace: Sovereign and financial balance sheets Doom Loops. *Review of Economic Studies* 85: 1781–823. [[CrossRef](#)]
- Froot, Kenneth A., and Jeremy C. Stein. 1998. Risk management, capital budgeting and capital structure policy for financial institutions: An integrated approach. *Journal of Financial Economics* 47: 55–82. [[CrossRef](#)]
- Gao, Pengjie, Paul Schultz, and Zhaogang Song. 2017. Liquidity in a Market for Unique Assets: Specified Pool and To-Be-Announced Trading in the Mortgage-Backed Securities Market. *The Journal of Finance* 72: 1119–70. [[CrossRef](#)]
- Garbade, Kenneth. 1999. *Fixed Income Analytics*. Cambridge: MIT Press.
- Gibson, Heather D., Stephen G. Hall, and George S. Tavlasi. 2017. A suggestion for constructing a large time-varying conditional covariance matrix. *Economics Letters* 156: 110–13. [[CrossRef](#)]
- Glosten, Lawrence R., and Paul R. Milgrom. 1985. Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics* 14: 71–100. [[CrossRef](#)]
- Hagströmer, Björn, Richard Henricsson, and Lars L. Nordén. 2016. Components of the Bid—Ask Spread and Variance: A Unified Approach. *Journal of Futures Markets* 36: 545–63. [[CrossRef](#)]
- Ho, Thomas, and Hans R. Stoll. 1981. Optimal Dealer Pricing under Transactions and Return Uncertainty. *Journal of Financial Economics* 9: 47–73. [[CrossRef](#)]
- Ho, Thomas, and Hans R. Stoll. 1983. The Dynamics of Dealer Markets under Competition. *Journal of Finance* 38: 1053–74. [[CrossRef](#)]
- Huang, Roger D., and Hans R. Stoll. 1997. The Components of the Bid-Ask Spread: A General Approach. *Review of Financial Studies* 10: 995–1034. [[CrossRef](#)]
- Koutmos, Gregory, and Andreas Pericli. 2000. Are multiple hedging instruments better than one? *Journal of Portfolio Management* 26: 63–70. [[CrossRef](#)]
- Leandro, Álvaro, and Jeromin Zettelmeyer. 2018. Evaluating Proposals for a Euro Area Safe Asset. Working Paper, Peterson Institute for International Economics, Washington, DC, USA.
- Lien, Donald, and Yiu Kuen Tse. 2011. Some recent developments in futures hedging. *Journal of Economic Surveys* 16: 359–96.
- Liu, Zijun, Yuliya Baranova, and Tamarah Shakir. 2017. Dealer Intermediation, Market Liquidity and the Impact of Regulatory Reform. Bank of England Staff Working Paper, Bank of England, London, UK, vol. 665.
- Lucas, André, Bernd Schwaab, and Xin Zhang. 2017. Modelling Financial Sector Tail Risk in the Euro Area. *Journal of Applied Econometrics* 32: 171–91. [[CrossRef](#)]
- Marcus, Alan J., and Evren Ors. 1996. Hedging corporate bond portfolios across the business cycle. *Journal of Fixed Income* 4: 56–60. [[CrossRef](#)]
- Morgan, Lawrence. 2008. Combination hedges applied to US markets. *Financial Analysts Journal* 64: 74–84. [[CrossRef](#)]
- Naik, Narayan Y., and Pradeep K. Yadav. 2003. Risk Management with Derivatives by Dealers and Market Quality in Government Bond Markets. *Journal of Finance* 58: 1873–904. [[CrossRef](#)]

- Schönbucher, Philipp J. 2003. *Credit Derivatives Pricing Models: Models, Pricing and Implementation*. London: Wiley.
- Stoll, Hans R. 1978. The Pricing of Security Dealer Services: An empirical study of Nasdaq stocks. *Journal of Finance* 33: 1153–72. [[CrossRef](#)]
- Stulz, René M. 1996. Rethinking risk management. *Journal of Applied Corporate Finance* 9: 8–24. [[CrossRef](#)]
- Tinic, Seha M. and Richard R. West. 1972. Competition and the Pricing of Dealer Services in the Over-the-Counter Market. *Journal of Financial and Quantitative Analysis* 7: 1707–27. [[CrossRef](#)]
- Yuan, Kathy. 2005. The Liquidity Service of Benchmark Securities. *Journal of the European Economic Association* 3: 1156–80. [[CrossRef](#)]



© 2019 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Article

Instantaneous Volatility Seasonality of High-Frequency Markets in Directional-Change Intrinsic Time

Vladimir Petrov ^{1,*}, Anton Golub ² and Richard Olsen ³

¹ Department of Banking and Finance, University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland

² Flov Technologies AG, Gotthardstrasse 26, 6300 Zug, Switzerland; anton.golub@flovtec.com

³ Lykke Corp., Alpenstrasse 9, 6300 Zug, Switzerland; richard.olsen@lykke.com

* Correspondence: vladimir.petrov@uzh.ch

† Current address: Department of Banking and Finance, University of Zurich, Plattenstrasse 14, 8032 Zurich, Switzerland.

Received: 28 February 2019; Accepted: 26 March 2019; Published: 1 April 2019

Abstract: We propose a novel intraday instantaneous volatility measure which utilises sequences of drawdowns and drawups non-equidistantly spaced in physical time as indicators of high-frequency activity of financial markets. The sequences are re-expressed in terms of directional-change intrinsic time which ticks only when the price curve changes the direction of its trend by a given relative value. We employ the proposed measure to uncover weekly volatility seasonality patterns of three Forex and one Bitcoin exchange rates, as well as a stock market index. We demonstrate the long memory of instantaneous volatility computed in directional-change intrinsic time. The provided volatility estimation method can be adapted as a universal multiscale risk-management tool independent of the discreteness and the type of analysed high-frequency data.

Keywords: instantaneous volatility; directional-change; seasonality; forex; bitcoin; S&P500; risk management; drawdown

1. Introduction

All events relevant to the performance of the financial system such as political decisions, natural disasters, or economic reports rarely happen synchronously and are typically not equally spaced in time. A sequence of them has a non-homogeneous nature and is not characterised by any vital autocorrelation function. Ultimately, the change of days and nights, as well as seasons, is dictated by the natural structure of the physical world which is barely connected to the flow of financial activity. Human minds, with the whole diversity of peculiar and indescribable characteristics, are primal engines of all market's evolutionary shifts. The global market, where the majority of transactions happen online and where traders, dealers, and market makers are distributed all around the world, is completely blind and deaf to the periodicity of days and nights, as well as to the climate factors of any standalone region of the Earth. New statistical tools, agnostic to the flow of the physical time, should be employed in order to handle the inner periodicity of the financial activity efficiently. In this work, we explore a concept of the endogenously defined time in finance applied to evaluate seasonality in markets' activity.

Probabilities of price drops and price rises between the running price maxima and running price minima are one of the most well-known risk-factors in finance. These probabilities are also called drawups and drawdowns. Numerous research works have focused on the analysis of the size, periodicity, and the time of recovery associated with drawups and drawdowns in traditional markets. The joint Laplace transform was utilised by Taylor (1975) for deriving the expected time until a new

drawup in a drifted Brownian motion occurred (traditionally considered as the model for historical price returns). The joint probability of observing a drawup of a given size after a drawdown, during a given term, was analysed as a homogeneous diffusion process in [Pospisil et al. \(2009\)](#). [Zhang \(2015\)](#) derived the joint probability in the context of exponential time horizons (the horizons are exponentially distributed random variables). The authors also described the law of occupation times for both drawup and drawdown processes. These and other theoretical findings connected to the price trend reversals were successfully applied to real financial problems such as studding market crashes.

Market crashes, pronounced in the abnormal price decreases, might severely impact the long-term stability of markets. It is especially important to estimate the probability of the next crash occurrence within a given period of time. Many research works were done on studying the crash probabilities using the normal distribution of price returns as the proxy for the real process. However, extreme price drops occur more often in the real world than what should happen when the distribution of returns coincides with the normal one. Fat-tailed distributions of returns ground the observed phenomenon. The distributions were discovered in the stock market ([Jondeau and Rockinger 2003](#); [Koning et al. 2018](#); [Rachev et al. 2005](#)), in the Forex (FX) ([Cotter 2005](#); [Dacorogna et al. 2001](#)), as well as in Bitcoin, prices ([Begušić et al. 2018](#); [Liu et al. 2017](#)). The fat tails, accompanied by the extensive discontinuity of the price curve (jumps), make the equally spaced time intervals inconvenient for high-frequency market analysis. Research tools, capable of working independently to the price distribution, should be called to deal with the erratic price evolution. Prices, at which drawdowns and drawups of the given size are registered, are independent of the time component of the price progression. Thus, the drawdown and drawups are the concepts especially useful of handling the dynamics of high-frequency markets. A sequence of drawdowns and drawups, following each other, can describe the evolution of a time series purely from the price point of view. The efficient set of forecasting techniques aimed at identifying appropriate conditions for future market crashes should inevitably be supplied by risk-management tools managing sequences of drawdowns and drawups.

In this research work, we investigate the connection between the observed number of alternating drawdowns and drawups (directional-change intrinsic time measure) and the instantaneous volatility. Non-parametric estimation of instantaneous volatility is still a relatively new topic which, to the extent of our knowledge, has not been studied before from the point of view of directional-change intrinsic time. Obtained in the work, analytical expressions are employed to reveal the seasonality structure of instantaneous volatility typical for high-frequency exchange rates. The described tools and experiments contribute to the collection of existing literature on directional-change intrinsic time and the seasonality properties of high-frequency markets. The tools will benefit high-frequency traders whose computer algorithms primarily operate on ultra-short time intervals where the short-term properties dominate over the long-term statistical characteristics ([Gençay et al. 2001](#); [Hasbrouck 2018](#)).

Three distinctive markets were considered in the work: FX (EUR/USD, EUR/JPY, and EUR/GBP), stocks (S&P500), and crypto (BTC/USD). All experiments are performed on the time series of the highest granularity: tick-by-tick data. That high granularity is essential considering the substantially growing interest in high-frequency trading after the 2008 financial crisis ([Kaya et al. 2016](#)). The data corresponds to the recent time period from 2011 to 2018 and is obtained from the largest trading venues (*JForex* and *Kraken*) opened for traders of any size. Each of the time series used in the empirical analysis is at least four years long. Such an extended length allows us to claim that properties specific for any particular period of time should not be pronounced in the obtained results.

The outline of the remaining paper is as follows. Section 2 provides a brief overview of research works on the properties of drawups and drawdowns. Section 3 gives detailed reasoning on the need for directional-change intrinsic time and describes a set of literature where the concept was successfully applied. Existing studies on the volatility seasonality of high-frequency markets is provided in Section 4. Section 5 describes the data used in the experiments and Section 6 outlines how the number of directional changes is connected to the instantaneous volatility. In Section 7 we present all results obtained by the traditional, as well as the novel, volatility measurement techniques and also describe

the application of theta time concept aimed to minimise the seasonality pattern. Section 8 concludes the main body of the paper and proposes the potential use of the developed technique. Appendix A concludes the paper by presenting a set of experiments where the comparison of considered markets seasonality patterns is presented.

2. Drawdowns and Drawups: An Introduction

Probabilities of financial drawdowns and drawups were extensively studied and presented in multiple seminal research works. Drawdowns of extensive size are usually associated with market crashes. Carr et al. (2011) proposed a new insurance technique aimed to protect investors against unexpected price moves. The authors also covered a novel way of hedging liabilities associated with these risks. Zhang and Hadjiliadis (2012) employed statistical properties of drawdowns as an estimate of the stock default risk and also provided a risk-management mechanism affecting the investor’s optimal cancellation timing. In Schuhmacher and Eling (2011) drawdowns are considered as one of 14 reward-to-risk ratios alternative to the widely known performance measures such as the Sharpe ratio. In Grossman and Zhou (1993) and Chekhlov et al. (2005) the properties of drawdowns were also applied as an estimate of the portfolio optimisation problem. The latter can be personalised to match traders’ or investors’ expectations and their tolerance to the size and the length of the market disruption.

Drawdowns D_t and drawups U_t , also called rallies in Hadjiliadis and Večeř (2006), registered by the moment of time t , depend on the running price maxima \bar{S}_t and the running price minima \underline{S}_t (Dassios and Lim 2018; Landriault et al. 2015; Mijatović and Pistorius 2012; Zhang and Hadjiliadis 2012). These reference points hinge on the set of historical prices S_s and are mathematically defined in the following way:

$$\bar{S}_t = \sup\{S_s : 0 \leq s \leq t\} \quad \text{and} \quad \underline{S}_t = \inf\{S_s : 0 \leq s \leq t\}, \tag{1}$$

where $t \geq 0$ and the interval $[0, t]$ is fixed. Drawdowns and drawups are the differences between the final price of the given time interval S_t and the registered local maxima and minima:

$$D_t = \bar{S}_t - S_t \quad \text{and} \quad U_t = S_t - \underline{S}_t. \tag{2}$$

The waiting time τ_a^D becomes measured once a price curve experiences a drawdown D_t of the size a . Similarly, τ_a^U is the waiting time associated with a drawup of the size a . In details:

$$\tau_a^D = \inf\{t \geq 0 : D_t \geq a\} \quad \text{and} \quad \tau_a^U = \inf\{t \geq 0 : U_t \geq a\}. \tag{3}$$

The waiting time τ_a measures the period of physical time which elapses before the first drawdown (potentially interpreted as a market crash) becomes registered.

3. Directional-Change Intrinsic Time

The existing literature on risk-management techniques primarily relies on physical time as a measure of the length and periodicity of financial events. In other words, the existence of a universal clock dictating the evolution of the market prices is assumed. However, the volatilities of different time resolutions behave differently (Müller et al. 1997). The volatility size depends on the scale of the entire time series as well as on the moment when the price activity started to be observed. More robust techniques which are beyond the limits of physical time are needed to handle this stochasticity.

The concept of directional-change intrinsic time (Guillaume et al. 1997) is one of the methods capable of replacing the universal physical clock with intrinsic one. This is an event-based framework which considers the activity of market prices as the indicator of the transition between its different states. The framework dissects a price curve into a collection of sections characterised by alternating trends of the arbitrary defined size. The essence of the concept is closely related to the meaning of

drawdowns and drawups: the collection of directional changes following each other can be interpreted as the alternating sequence of drawdowns and drawups. The frequency of price changes in physical time does not play any role in the directional change dissection procedure.

The concept of trend directional changes provided by [Guillaume et al. \(1997\)](#) is capable of connecting the continuous flow of physical time with the endogenous evolution of price returns. According to the event-based space proposed by Guillaume, only a sequence of price trends continuously alternating in direction has to be considered. The price curve gets dissected into a collection of alternating drawups and drawdowns or trend rises and trend falls correspondingly. Each elementary trend ends once a new price curve reversal is observed. Continuous price moves towards the direction of the latest trend change are called overshoots. The current state of the system changes only at the moments when the trend of the given size reverses its direction. Thus, the set of intrinsic events is decoupled from the flow of physical time. Instead, it depends only on the size of considered drawups and drawdowns labelled by the threshold δ . An example of a price curve dissected into a collection of directional changes is provided in [Figure 1](#).

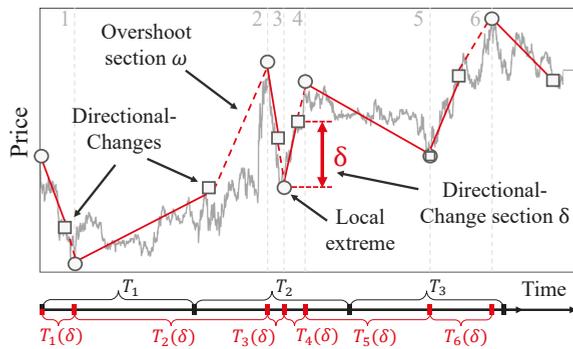


Figure 1. A part of EUR/USD price curve (grey) dissected into a set of directional-changes (grey squares) using a directional-change threshold δ . The size of the arbitrary chosen threshold is presented in the middle of the figure. Grey circles mark local extremes between two consecutive directional changes. The vertical distance between each directional-change and preceding extreme price is bigger or equal to the size of the threshold δ . Vertical dashed lines indicate the end of each trend section (identified only after the next event becomes observed) and go through the local extremes (circles). The timeline below the plot contains equal time intervals T_1, T_2, T_3 and length of each directional-change section $T_1(\delta), \dots, T_6(\delta)$.

The density of directional-change intrinsic events depends only on the price curve evolution and the considered trend size. The stochastic nature of price evolution results in the phenomenon depicted in [Figure 1](#): non-equal number of intrinsic events (empty squares) correspond to the equal periods of physical time. The physical interval T_1 contains only the end of the Section 1 (sections coincide with the intrinsic events and are separated by the dashed vertical lines) while the equal interval T_2 hosts three segments, namely 2, 3, 4. This property of directional-change intrinsic time can be engaged as the efficient noise filtering technique: the intrinsic time ignores price changes between directional change. At the same time, it allows us to efficiently capture the most relevant to risk management information: precise moments of all trend changes. The equally spaced time intervals typically employed in the financial analysis are not capable of doing anything of the above: price timestamps, evenly spaced through periods T_1, T_2 , and T_3 , do not contain information on the extreme price curve activity located in the period T_2 . This disability of the traditional price analysis techniques over stochasticity of the market's activity develops into volatility estimators that are too stiff and biased.

The concept of directional-change intrinsic time, applied for studying historical price, returns reveals multiple statistical properties of high-frequency markets. [Guillaume et al. \(1997\)](#) were the first researchers to uncover a scaling law¹ relating the expected number of directional-changes $N(\delta)$ observed over the fixed period to the size of the threshold δ . Mathematically:

$$N(\delta) = \left(\frac{\delta}{C_{N,DC}} \right)^{E_{N,DC}}, \quad (4)$$

where $C_{N,DC}$ and $E_{N,DC}$ are the scaling law coefficients. [Glattfelder et al. \(2011\)](#) employed the directional-change framework to discover 12 independent scaling laws which hold across three orders of magnitude and are present in 13 currency exchange rates. Later [Golub et al. \(2017\)](#) described a successful trading strategy exploiting a collection of tools build upon directional-change intrinsic time. The proposed strategy is characterised by the annual Sharpe ratio greater than 3.0. The persistence of revealed scaling laws became the base elements for the tools designed to monitor market’s liquidity at multiple scales ([Golub et al. 2014](#)).

4. Seasonality

4.1. Traditional Markets

The returns seasonality is the well-known statistical characteristic of developed markets such as FX and stocks. [Rozeff and Kinney \(1976\)](#) studied the comprehensive set of historical stock data which spans from 1904 to 1974 and found a higher mean of return in the January distribution of returns compared with most other months. They also underline noticeably high mean returns in July, November, and December, and low mean returns in February and June. [Gultekin and Gultekin \(1983\)](#) empirically examined stock market seasonality in major industrialised countries. They aimed at investigating the existence and the shape of the stock market seasonality pattern in foreign securities markets. The confirmed seasonal patterns in the stock returns supplied the further understanding of the seasonality anomaly. [Seif et al. \(2017\)](#) studied seasonal anomalies in advanced emerging stock markets and provides re-examination of the markets efficiency. The authors did not find the confirmation of the January effect but confirmed the month of the years, the day of the week, the holiday, and the week of the year effects. The recent work [Fang et al. \(2018\)](#) presents a strong link between school holidays and market returns across 47 countries. The authors demonstrate that the returns in the month after major school holidays are 0.6% to 1% lower than at other times. The provided evidence states that post-school holiday small returns are explained by the investors’ inattention during these periods. The reduced attention results in news effects being incorporated noticeably slower into prices than within the active trading periods. We also underline the relevance of other research works on the stock market seasonality: [De Bondt and Thaler \(1987\)](#); [Keim \(1983\)](#); [Zarowin \(1990\)](#), among others.

Daily, weekly, and annually seasonality patterns are also inevitable components of the set of FX stylised facts. [Müller et al. \(1990\)](#) analysed four foreign exchange spot rates against the USD over three years. Authors’ intra-day and intra-week analysis show that there are systematic variations of volatility present even within business hours. They also discovered daily and weekly patterns for the average bid-ask spread. [Dacorogna et al. \(1993\)](#) studied daily and weekly FX seasonality patterns from the geographically distributed trading point of view. The trading activity divided in three general components (East Asia, Europe, and America) was approximated by a polynomial activity function during business hours. The combined model was closely fitted into the empirical volatility (activity) seasonality data. The authors found that strongly seasonal activity autocorrelation can be approximated by the hyperbolic function. [Bollerslev and Domowitz \(1993\)](#) examined behaviour of quote arrivals and bid-ask spreads for continuously recorded deutsche mark-dollar exchange rate data

¹ A basic polynomial functional relationship where a change in input results in a proportional change in output.

over time, across locations, and by market participant. The authors find the relation of the considered information to the seasonality patterns typically observed in the deutsche mark-dollar exchange rate. [Ito and Hashimoto \(2006\)](#) showed U-shaped intra-day activities of deals and price changes as well as return volatility for Tokyo and London participants of USD/JPY and EUR/USD markets. The authors also note that the U-shape was not found for New York participants. A set of well-known seasonality factors was confirmed: the high activities at the opening of the markets, high correlations between quote entries and deals, and higher trading activities associated with narrow spreads.

The seasonality of the rapidly evolving cryptocurrency domain is still insufficiently studied in the financial world. The next section unwraps some of the facts about cryptocurrencies and presents outcomes of the previous studies which have to be considered in the current work.

4.2. Bitcoin Seasonality

Bitcoin is the first successful pioneer in the crypto domain. It is also the most famous representative of the cryptocurrency markets. Bitcoin was described by [Nakamoto \(2008\)](#) and created in 2008 as an alternative to the classical financial system. The cryptocurrency was rapidly tagged as the “peer-to-peer version of electronic cash”. Bitcoin and its underlying technology, blockchain², swiftly gained attention from the technologically savvy community and media. The peer-to-peer payment systems soon became one of the most debatable topics at all levels of modern society. Over a thousand alternative cryptocurrencies, based on the similar cryptography concept ([Roy and Venkateswaran 2014](#)), emerged since the time Bitcoin was invented. Some part of them became accessible for trading at various electronic venues also known as crypto-exchanges³.

In contrast to the traditional FX market, cryptocurrency trades happen 24 h per day and seven days per week, independently of the holidays and seasons. Additionally, cryptocurrency trading activity is more uniformly distributed across the globe. There are not many big geographically segregated financial institutions where trades happen according to the working schedule (in contrast to the global FX trading centres described in [Dacorogna et al. \(1993\)](#)). There is still the limited acceptance of this new financial instrument by international organisations with access to sizable funds⁴. As result, the seasonality patterns prevalent in the world of cryptocurrencies could be incomparable with the ones typical for the FX or stock markets. High volatility has been one of the most pronounced characteristics of the Bitcoin markets (see, for example, [Dyhrberg \(2016\)](#)). Bitcoin’s trend drastically changes and their persistence indicates the aggregated expectations and trading actions of all market participants. Unstable trends also reveals the Bitcoin price sensitivity to exogenous stress factors. The high scale of Bitcoin trend changes attracts researchers to employ modern technologies in order to foresee the future price dynamics ([Shintate and Pichl 2019](#)).

There are a few research works concerned with the statistical properties of cryptocurrency markets. [Sapuric and Kokkinaki \(2014\)](#) analysed realised volatility of Bitcoin returns within a 4-year time interval to understand what the prime characteristics of its price activity are. They confirmed Bitcoin’s high volatility, but emphasised that traded volume should be taken into account when computing the precise value⁵. The authors compared Bitcoin with conventional financial instruments, including gold, and several national currencies. They demonstrated that the calculated volatility significantly decreases when the traded volumes are included in the model.

[Haferkorn and Diaz \(2014\)](#) studied seasonality patterns of the number of payments performed in three cryptocurrencies: Bitcoin (classified as a worldwide payment system), Litecoin (open source

² A growing list of records containing information on the ownership of all existing Bitcoins.

³ Information on all cryptocurrencies and trading venues can be found at [Coinmarketcap.com](https://coinmarketcap.com).

⁴ At the moment of writing the paper, Wall Street and other big financial hubs are considering trading cryptocurrencies, which will potentially result in the higher segregation level.

⁵ According to the Bank for International Settlements the daily average FX trading volume in April 2016 was \$5.1 trillion ([BIS 2016](#)) when the highest registered volume in the crypto market is to the date only \$45.8 billion (<https://coinmarketcap.com/charts/>).

software project), and Namecoin (decentralised name system). Their research confirmed that the monthly or yearly seasonality is not typical for the crypto market. The only robust weekly pattern was found in Bitcoin prices. Litecoin and Namecoin had weak or no patterns at all. The authors state that there is also no significant correlation between the returns of observed exchange rates. Authors speculate about the reason of this phenomenon. They say that these cryptocurrencies have similar core architecture, but they all have been created to serve specific needs.

de Vries and Aalborg (2017) made another attempt to discover Bitcoin seasonality patterns. They analysed daily traded volume, daily transaction volume, and Google trends (the number of searches for the word “bitcoin”). The author also inspected the seasonality of the number of transactions performed from individual blockchain accounts. All of the measurements demonstrated no particular periodicity.

Eross et al. (2017) gave a more positive answer on the existence of the most famous cryptocurrency intraday seasonality. The authors investigated Bitcoin returns, volume, realised volatility, and bid-ask spreads to reveal several intraday stylised facts. A significant negative correlation was found between returns and volatility. Volume and volatility were shown to have a considerably positive correlation. The authors attribute such patterns to the European and North American traders, as well as the insufficient number of market makers in the whole crypto space.

The seasonal heteroscedasticity affects the results of statistical studies of intraday and intra-week price properties. The rapidly evolving electronic high-frequency trading is highly affected by the exchange rate seasonality properties too (Cont 2011b). Therefore, the returns’ seasonality has to be treated at the first priority.

More information on order patterns in time series can be found in the work Bandt and Shiha (2007). The authors determine probabilities of order patterns in Gaussian and autoregressive moving-average processes, which can be directly applied to the financial time series analysis.

5. Data

Three FX exchange rates were used in the work: EUR/USD, EUR/JPY, and EUR/GBP. The covered time interval is from January 2011 to January 2016 and includes 109,069,357, 134,737,397, and 88,704,676 ticks correspondingly. The source of the data is the *JForex* trading platform developed by the Swiss bank and marketplace *Ducascopy*, which provides various types of market data in the highest resolution⁶.

Bitcoin price changes observed at the *Kraken* crypto-exchange were downloaded from the *Bitcoincharts* online platform supplying financial and technical data related to the Bitcoin network⁷. The studied time interval is from January 2014 to April 2018 and includes 4,778,429 ticks.

The stocks market is represented in this work by the index that currently comprises 505 common stocks listed by 500 large-cap companies on the US stock market: S&P 500 (with the ticker SPX500). The high-frequency data has been downloaded from the *JForex* platform too. The selected dataset spans from January 2012 to January 2017. The total number of ticks is 38,931,943.

There is a substantial difference in the number of ticks in time series used in the work. The difference reflects the distinct activity in the selected traditional and novel markets. That discrepancy does not undermine the further results since our goal is to apply the novel volatility measurement techniques, which is agnostic to the number of ticks per period of time (see Section 3). Moreover, the discrete price impact on the instantaneous volatility will be described in Section 7.3. The same is correct for the slight time-span shifts. The compatibility of the results is not affected: the designed experiments aim to depict the statistical properties of the volatility seasonality and not to compare price behaviour at any particular historical moment of time.

⁶ <https://www.ducascopy.com/swiss/english/forex/jforex/>.

⁷ <http://api.bitcoincharts.com/v1/csv/>.

The fully functional code used in the project can be downloaded from the author's GitHub repository⁸.

Inner Price

Any collection of historical prices typically assumes two values: the best bid (buy) and the best ask (sell). These prices are collected from the complete order books specific for the given exchange. The order books contain all clients orders submitted in the market at the given moment. The non-zero price difference between the best offers on the sell and buy sides, called spread, indicates the level of a market's liquidity (Bessembinder 1994; Dyhrberg et al. 2018; Menkhoff et al. 2012). It also has a direct connection to realised volatility (Bollerslev and Melvin 1994), and is an indicator of the transaction cost (Hartmann 1999). Another role of the spread is to show the extent of uncertainty the market has on the fair price of the traded asset. The size of the spread constantly changes over time, together with the level of uncertainty. This fact does not allow us to employ only bid or ask prices to study properties of the whole market at the micro level since some part of the information is in the risk to be lost while analysing intraday data. The average of these two values (mid-price) is also not the best alternative since it does not keep the knowledge on the size of the current spread. Therefore, an alternative measure should be chosen to apply the directional-change algorithm to the real data analysis.

The trend-dependent concept of inner price was selected to resolve the spread issue. Inner price, specific for the given moment of time, is defined as the bid or the ask price depending on the direction of the current trend. In other words, it is the price where the spread is deduced. The following example demonstrates details of the concept. According to the directional-change algorithm, one should wait for the price increase by δ percents from the local price minimum to register a new directional change if the current trend is downward. In this case, the value of the inner price coincides with the best price on the offer side of the order book, that is, the ask price. A new intrinsic event will tick only when the distance between the latest bid price and the inner price reaches the size of the chosen directional-change threshold δ . Alternatively, the inner price takes the value of the best bid price, and the distance is measured between the newest ask and the extreme if the current mode is upward.

6. Methods

Theoretical researchers mostly rely on the Brownian motion as the proxy for price returns of real financial markets (one of the most famous examples is the work of Black and Scholes (1973)). The analogy between historical price moves and changing coordinates of an ensemble of molecules in thermodynamics is the motive behind this common approach. Osborne (1959) shows in the classical work that the steady-state distribution of log-returns in the stock market is the probability distribution for a particle in Brownian motion. It is important to emphasise that at the telegraph-driven time of Osborne's publication (1959) the structure and the dynamic of the market was very different from the ones typical for our modern digital world. The present-day trading has almost completely moved to the digital online space instead of the physical trading floors where all deals happened more than a half a century ago (see Harris (2003) for the historical endeavour on the evolution of trading and exchanges). In this space, a signal can easily propagate through international borders with the speed of light. It stipulated the majority of trades to happen in a fully automated way. The significantly bigger diversity of the market participants and the ability of the high-frequency trading made the financial stylised facts more important today than ever. The stylised facts describe the deviations of real price returns from the theoretical Brownian motion (see Cont (2001a) for the set of stylised empirical facts). Nevertheless, in our work, we will operate with Brownian motion as the core for the analytical part of the research.

⁸ <https://github.com/VladUZH/VIПетров>.

The choice of Brownian motion, employed in this work, is justified by two reasons. First, the statistical properties of the number of directional-change intrinsic events, studied in this work, is agnostic to the flow of physical time. Second, the divergence of empirical results from the properties of the selected model, if any, helps in understanding the features of the real markets better.

6.1. *Waiting Time*

We model the set of prices $\{S_t: t \geq 0\}$ as an arithmetic Brownian motion with trend μ and volatility σ :

$$dS_t = \mu dt + \sigma dB_t. \tag{5}$$

In terms of the directional-change intrinsic time framework, $T_{up}(\delta_{up})$ denotes the time for an upward directional change of the size $\delta_{up} > 0$ to unfold. In other words, it is the time interval which passes until the price increases by δ_{up} percents from the local minimum m_t . Technically:

$$T_{up}(\delta_{up}) = \inf\{t > 0 : \frac{S_t - m_t}{m_t} \geq \delta_{up}\}, \tag{6}$$

where

$$m_t := \inf_{\epsilon \in [0,t]} S_\epsilon. \tag{7}$$

Similarly, $T_{down}(\delta_{down})$ is the time of a downward directional change of the size $\delta_{down} > 0$:

$$T_{down}(\delta_{down}) = \inf\{t > 0 : \frac{M_t - S_t}{M_t} \geq \delta_{down}\}, \tag{8}$$

where

$$M_t := \sup_{\epsilon \in [0,t]} S_\epsilon. \tag{9}$$

Both of these equations are also known in the literature as waiting times of drawups and drawdowns (see Section 1). It is shown in Landriault et al. (2015) that expected times of a drawup δ_{up} and a drawdown δ_{down} depend on the volatility and the trend of the drifted Brownian motion. It can be mathematically expressed as

$$\mathbb{E}[T_{up}(\delta_{up})] = \frac{e^{-\frac{2\mu}{\sigma^2}\delta_{up}} + \frac{2\mu}{\sigma^2}\delta_{up} - 1}{\frac{2\mu^2}{\sigma^2}}, \tag{10}$$

and

$$\mathbb{E}[T_{down}(\delta_{down})] = \frac{e^{\frac{2\mu}{\sigma^2}\delta_{down}} - \frac{2\mu}{\sigma^2}\delta_{down} - 1}{\frac{2\mu^2}{\sigma^2}}. \tag{11}$$

Using the Taylor expansion $e^{\pm \frac{2\mu}{\sigma^2}\delta} = 1 \pm \frac{2\mu}{\sigma^2}\delta + \frac{(\frac{2\mu}{\sigma^2}\delta)^2}{2!} + \mathcal{O}(\mu^3)$ and letting $\mu \rightarrow 0$, one can recover that in the case with no trend the equation simplifies to

$$\mathbb{E}[T_{up}(\delta)] = \mathbb{E}[T_{down}(\delta)] = \frac{\delta^2}{\sigma^2}. \tag{12}$$

These equations establish a scaling law dependence between waiting times of a directional change, volatility, and the selected size of the directional-change threshold. Indeed, in the analysis of Glattfelder et al. (2011) it was empirically found that in the FX market the average waiting time is proportional to the second power of the directional-change threshold δ used to identify alternating trends:

$$\langle T(\delta) \rangle \sim \delta^2. \tag{13}$$

The closeness of Equations (12) and (13) confirms the assumption that the evolution of high-frequency prices expressed in terms of the directional-change intrinsic time has similar properties to the random walk.

6.2. Number of Directional Changes

Let $N(\delta_{down}; \sigma, \mu, [0, T])$ denote the number of drawdowns of the size δ_{down} observed within the time interval $[0, T]$ in Brownian motion process with parameters μ and σ . Since the sequence $T_{down}(\delta_{down})_1, T_{down}(\delta_{down})_2, \dots$ is the sequence of non-negative, independent, and identically distributed random variables, the sequence $\{\psi_n; n \in \mathbb{N}\}$ where $\psi_n = T_{down}(\delta_{down})_1 + \dots + T_{down}(\delta_{down})_n + \dots$ is the renewal point process. Thus, $N(\delta_{down}; \sigma, \mu, [0, T])$ can be considered as the renewal counting process and its values can be found applying the Theorem 6.1.1 of [Rolski et al. \(2009\)](#) ([Landriault et al. 2015](#)) to the waiting time Equation (11):

$$\lim_{T \rightarrow \infty} N(\delta_{down}; \sigma, \mu, [0, T]) = \mathbb{E}[T_{down}(\delta_{down})]^{-1} T = \frac{T \frac{2\mu^2}{\sigma^2}}{e^{\frac{2\mu}{\sigma^2} \delta_{down}} - \frac{2\mu}{\sigma^2} \delta_{down} - 1}. \tag{14}$$

Correspondingly, the expected number of drawups $N(\delta_{up}; \sigma, \mu, [0, T])$ takes the form

$$\lim_{T \rightarrow \infty} N(\delta_{up}; \sigma, \mu, [0, T]) = \mathbb{E}[T_{up}(\delta_{up})]^{-1} T = \frac{T \frac{2\mu^2}{\sigma^2}}{e^{-\frac{2\mu}{\sigma^2} \delta_{up}} + \frac{2\mu}{\sigma^2} \delta_{up} - 1}. \tag{15}$$

Equations (14) and (15), combined together, give the estimate of the number of directional changes consequently following each other:

$$\mathbb{E}[N(\delta_{up}, \delta_{down}; \mu, \sigma, [0, T])] = \frac{2T \frac{2\mu^2}{\sigma^2}}{e^{-\frac{2\mu}{\sigma^2} \delta_{up}} + e^{\frac{2\mu}{\sigma^2} \delta_{down}} + \frac{2\mu}{\sigma^2} (\delta_{up} - \delta_{down}) - 2}. \tag{16}$$

The expression is simplified in the trend-less case ($\mu \rightarrow 0$) to the following form:

$$\mathbb{E}[N(\delta_{up}, \delta_{down}; \sigma, [0, T])] = \frac{2T\sigma^2}{\delta_{up}^2 + \delta_{down}^2}. \tag{17}$$

The theoretical dependence of the number of directional changes and the properties of underlying process resemble the empirical observations of [Guillaume et al. \(1997\)](#). The authors mention there that $N(\delta) \sim \delta^{-2}$ (for $\delta = \delta_{up} = \delta_{down}$).

Monte Carlo statistical tests were performed to numerically verify the accuracy of Equations (10), (11) and (16). Results of the tests are provided in Table 1. We selected only positive trend values μ since the equations are symmetrical with respect to the direction of the trend. Values in Table 1 exhibit high similarity of both empirical and theoretical results.

The meaning behind the provided equations is that the absolute size and the ratio of directional-change thresholds used to dissect a price curve into a sequence of upward and downward trends affect the frequency and the total number of events registered within a given time interval. It follows from Equations (14) and (15) that the combination $\gamma = \frac{\mu}{\sigma^2}$ is the crucial factor affecting the expected number of intrinsic events⁹. We check the number of directional changes registered by a couple of thresholds in three extreme scenarios: $\frac{\mu}{\sigma^2} = 0$ (Figure 2a), $\frac{\mu}{\sigma^2} \ll 0$ (Figure 2b), and $\frac{\mu}{\sigma^2} \gg 0$

⁹ The expression γ is known in the insurance industry as “adjustment coefficient” or “the Lundberg exponent” ([Asmussen and Albrecher 2010](#)). It finds its application in the ruin theory dating back to 1909 ([Lundberg 1909](#)). It is also described as the optimal information theoretical betting size called Kelly Criterion ([Kelly 2011](#)).

(Figure 2c). A diverse set of dissection procedures was applied to the randomly generated time series defined by the parameters γ . All results were composed as a heatmap where each point corresponds to the number of directional changes observed by a pair of thresholds $\{\delta_{up}, \delta_{down}\}$ (Y- and X-axis of the plots) in a time series of the given length (Figure 2).

Table 1. Waiting times and number of directional changes in a Monte Carlo simulation. μ and σ are parameters of the Brownian motion used for the test. There are 10^9 ticks in the simulated time series. N_{DC}^{MC} , $\langle T_{up}^{MC} \rangle$, and $\langle T_{down}^{MC} \rangle$ are the numbers of directional changes and the average waiting times registered in the Monte Carlo simulation. $\mathbb{E}[N_{DC}]$, $\mathbb{E}[T_{up}]$, and $\mathbb{E}[T_{down}]$ are theoretical values dictated by Equations (16), (10) and (11) correspondingly. Values $\sigma_{T_{up}^{MC}}$ and $\sigma_{T_{down}^{MC}}$ are standard deviations of empirical and theoretical waiting times.

$\mu, \%$	$\sigma, \%$	$N_{DC}^{MC}/\mathbb{E}[N_{DC}]$	$\langle T_{up}^{MC} \rangle/\mathbb{E}[T_{up}]$	$\sigma_{T_{up}^{MC}}$	$\langle T_{down}^{MC} \rangle/\mathbb{E}[T_{down}]$	$\sigma_{T_{down}^{MC}}$
1	10	1.028	0.968	2.54×10^{-5}	1.019	2.53×10^{-6}
	20	1.009	0.989	2.78×10^{-6}	1.012	3.32×10^{-7}
	30	1.001	0.995	8.79×10^{-7}	1.033	9.58×10^{-8}
6	10	1.021	0.971	2.29×10^{-5}	1.043	2.59×10^{-6}
	20	1.005	0.993	2.94×10^{-6}	1.019	3.29×10^{-7}
	30	0.987	1.011	8.84×10^{-7}	1.034	9.98×10^{-8}
11	10	1.029	0.968	2.20×10^{-5}	1.011	2.78×10^{-6}
	20	0.994	1.006	2.72×10^{-6}	0.997	3.30×10^{-7}
	30	0.986	1.014	8.82×10^{-7}	1.017	1.02×10^{-7}

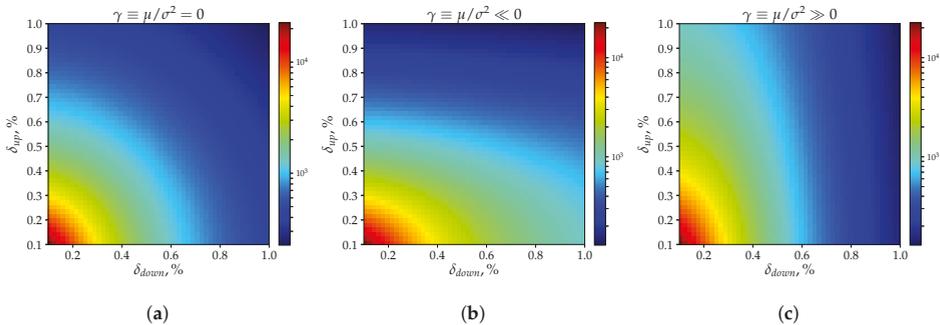


Figure 2. Heatmaps of the number of directional changes observed by the pair of directional-change thresholds $\{\delta_{up}, \delta_{down}\}$ (Y- and X-axis of the plots) in a timeseries of the given length (Geometrical Brownian Motion (GBM), 10^9 steps in each simulation). Selected trend and volatility values: (a) $\mu = 0$, $\sigma = 0.15$; (b) $\mu = -3$, $\sigma = 0.15$; (c) $\mu = 3$, $\sigma = 0.15$. The values on the plots coincide with the ones computed using Equation (17).

Panel 2a in Figure 2 corresponds to the set of experiments where the Brownian Motion trend is equal to zero. It follows from Equation (17) that in such conditions the value $\mathbb{E}[N(\delta_{up}, \delta_{down}; \sigma, [0, T])]$ should be constant along circular contours $\delta_{up}^2 + \delta_{down}^2 = \delta^2$ for $\delta > 0$. The colour gradient in the provided picture confirms the noted dependence. It is shown in panels 2b and 2c of Figure 2 that the circular contours transform into ellipses when the “adjustment coefficient” γ is significantly smaller or significantly bigger than zero. This phenomenon can be interpreted in the following way: if $\mathbb{E}[N(\delta_{up} = \delta_{down}; \gamma = 0, [0, T])]$ is the expected number of directional changes registered in the drift-less time series of given length and characterised by the fixed σ then for any γ greater or smaller than zero there is always such a couple of non equal thresholds $\{\delta_{up}, \delta_{down} | \delta_{up} \neq \delta_{down}\}$ that

$$\mathbb{E}[N(\delta_{up}, \delta_{down} | \delta_{up} \neq \delta_{down}; \gamma \neq 0, [0, T])] = \mathbb{E}[N(\delta_{up}, \delta_{down} | \delta_{up} = \delta_{down}; \gamma = 0, [0, T])]. \quad (18)$$

In other words, any process characterised by a certain degree of persistent trend could be treated as the one without the trend by tuning the size and the ratio of selected directional-change thresholds. The property is essential for risk management techniques constructed on top of directional-change intrinsic time approach. An example of real application of this fact is provided in [Golub et al. \(2017\)](#). The authors employed asymmetric thresholds to design an optimal inventory control function sensitive to the significant price trend changes.

6.3. Instantaneous Volatility

The volatility size of the financial time series is an inevitable component of any financial risk analysis. Therefore, a clear understanding of the way how volatility changes over time is particularly important for risk management and inventory control problems. Classical volatility estimation methods, also called “natural” or “traditional” estimators ([Cho and Frees 1988](#))¹⁰, primarily rely on physical time as the persistent measure of the intervals when the price returns should be computed. The fact that the variance of returns on assets tends to change over time creates obstacles on the way of employing the “traditional” volatility estimators. The changing variance, also known as the stochastic volatility, became a cornerstone for multiple research works (for example, [Ai and Kimmel \(2007\)](#); [Andersen and Lund \(1997\)](#); [Barndorff-Nielsen and Shephard \(2002\)](#); [Campbell et al. \(2018\)](#) and many others).

Values, computed by “natural” estimators, dominantly correspond to the integrated volatility of the studied process. The integrated volatility describes the averaged price activity over non-zero time intervals. Alternative estimators, designed to reveal the size of the volatility as the time interval approaches zero (instantaneous volatility), are mostly based on Fourier analysis¹¹ and require extensive computation efforts (see Chapter 3 in [Mancino et al. \(2017\)](#)). Therefore, new methods, capable of describing the price evolution independently of the flow of the price in physical time, should be employed to overcome the existing volatility estimation difficulties.

The directional-change intrinsic time concept is by design agnostic to the speed of the price change. Risk-management tools, based on top of the concept, automatically adapt their performance to treat the changing price activity better. This property of directional-change intrinsic time, together with analytical Equations (16) and (17), bring the idea of a new volatility estimator devoid of the shortcomings of the equidistant time in finance. It follows from Equation (17) that the volatility can be estimated for a trendless time series by counting the number of directional changes within the time interval $[0, T]$:

$$\sigma_{DC} = \delta \sqrt{\frac{N(\delta)}{T}}. \tag{19}$$

We use the superscript *DC* to distinguish the volatility computed through the directional-change intrinsic time from volatility computed by the traditional estimators. The latter we will mark by σ_{trad} .

Equation (19) solely computes the volatility part σ of the Brownian proces. That contrasts the “natural” volatility estimation techniques where the entire stochastic σdW_t part is typically measured. That stochastic factor includes the noise component dW_t . Therefore, the directional change approach employed for volatility measurements can be classified as the true estimator of the instantaneous volatility. Further, we apply Equation (19) to study changing dynamic of financial time series throughout one week. We reveal volatility seasonality patterns of three FX exchange rates, crypto market BTC/USD, and the stock index S&P500.

¹⁰ The work [Cho and Frees \(1988\)](#) is particularly interesting due to the analysis the authors did to compare volatilities computed by “natural” and “temporal” estimators. The latter employs time intervals measured between consequent and alternating price moves of fixed relative size and thus is very close to the approach presented in the current paper.

¹¹ The type of mathematical analysis applied to identify patterns or cycles in a normalised time series data.

7. Results

Empirical properties of three distinct markets will be discussed in this section. We omit the EUR/JPY, and EUR/GBP exchange rates in some of our experiments due to the fact that their properties are greatly similar to the properties of the most traded FX rate EUR/USD. In this case, EUR/USD is selected as the representative exchange rate of the entire FX domain.

7.1. Number of Directional Changes

Equations (16) and (17) connect the expected number of directional changes with parameters of the underlying Brownian motion process. The evolution of real historical returns have properties similar to the Brownian motion. The evolution sometimes compared to the sequence of the free particle moves (see Section 6). Thus, similar counters shown in Figure 2 should be found in heatmaps depicting the number of directional changes empirically registered in real data conditional that the assumption of the normal distribution of real returns is true. EUR/USD, BTC/USD, and SPX500 exchange rates were taken to verify the statement by replicating the same experiment done with the Brownian motion before (Figure 3). A collection of 40 directional-change thresholds ranging from 0.1% to 4.1% defines the scale of the heatmap grid. Colour schemes, used for the plots, have different scales due to the significantly bigger number of directional changes per a period of time in the BTC/USD case. Yellow solid lines indicate the examples of the areas where the number of direction changes is constant. The selected for the examples deltas are $\delta_{up} = \delta_{down} = \{1.15\%, 2.8\%\}$ (EUR/USD, Figure 3a), $\delta_{up} = \delta_{down} = \{1.4\%, 3.0\%\}$ (BTC/USD, Figure 3b), and $\delta_{up} = \delta_{down} = \{1.1\%, 2.9\%\}$ (SPX500, Figure 3c).

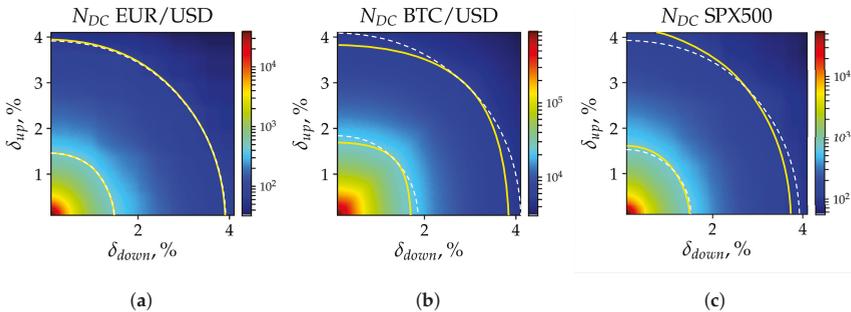


Figure 3. Heat map of the number of directional changes calculated in (a) EUR/USD, (b) BTC/USD, and (c) SPX500 time series. Each point on the grid represents the number of directional changes registered by a unique pair of thresholds $\{\delta_{up}, \delta_{down}\}$. Heatmaps have different scales. Yellow solid lines, specific for each heatmap, label the examples of the areas along which the number of intrinsic events is constant. The dashed lines represent the theoretical areas of the equal number of intrinsic events observed in case of the trend-less time series. White dashed lines are parts of circles centred around the left bottom corner of each picture. The lines go through the intersection of the solid yellow lines and the diagonal of each picture.

Curves in Figure 3a have an almost circular shape and are only slightly shifted towards the bigger δ_{down} values. This shift is present due to the downward trend experienced by the exchange rate from 2011 to 2016 (from \$1.4 to \$1.1 per EUR). BTC/USD exchange rate was much more unstable considering that the EUR/USD time series exhibited relative stability with no noticeable regime switches apart from the slow constant price depreciation. The price of Bitcoin grew with accelerating pace by more than 20 times in the second half of 2017 and then lost nearly 70% of its value at the beginning of 2018¹².

¹² It had a minimum at \$230 per Bitcoin, temporary maximum at about \$20,000, and then a drop to \$6000.

These significant trend changes are pronounced in Figure 3b by yellow contours notably deviated from the circular shape. The price roller-coaster caused considerable disparity of the number of registered directional changes and the ones predicted by Equation (17) (relevant to the trend-less case). As result, the solid price curves can be decomposed into two parts of independent ellipses similar to the ones observed for Brownian motion with non-zero “adjustment coefficient” γ (Figure 2b,c).

7.2. *Realised versus Instantaneous Volatility*

In the second experiment, we compared the annualised volatility computed by the traditional method (Equation (20)) and the volatility based on the observed number of directional changes (Equation (19)).

Returns R_t are defined as logarithms of the price change between S_t and S_{t-1} measured over equal periods of time. The number of returns n depends on the selected time interval Δt and equal to $n = T/\Delta t$ where T is the length of the entire tick-by-tick sample. Thus, the length of a sample can be computed ex-ante.

The whole set of returns was used to find the standard deviation of the time series. The measure is also known as realised volatility σ_{trad} :

$$R_t = \ln(S_t/S_{t-1}), \quad R_{avg} = \frac{\sum_{t=1}^n R_t}{n}, \quad \sigma_{trad} = \sqrt{\frac{\sum_{t=1}^n (R_t - R_{avg})^2}{n - 1}}. \tag{20}$$

The directional-change method does not define the number of observations ex-ante in contrast to the traditional approach. According to Equation (19), the size of the directional-change threshold δ determines only the expected number of measures (or timestamps) in the data sample of the given length. It is also worth saying that the price moves of the highest frequency, tick-by-tick, do not appear over any predefined period. They occur together with the flow of new orders in the market. The flow, initiated by thousands of independent traders’ demands, not synchronised with any periodical process. Thus, the time distance between two consecutive ticks can be represented by a fraction of a second as well as by several minutes. The equally spaced timestamps used to calculate returns for the “natural” estimators have a high chance to happen not at the moment of a new price change. The additional decision should be made on whether the historical price located before the timestamp or right after it should be selected to compute the corresponding return. The directional-change intrinsic time, in turn, directly reacts to the changes of the price levels. This flexibility of the intrinsic time makes it possible to use the data of the highest frequency: tick-by-tick prices.

Specifications of tools used to estimate volatility can affect the experiments results (Müller et al. 1997). Four increasing time intervals Δt_k , where $k = \{1, 2, 3, 4\}$, were selected to define the distance between each pair of consecutive prices S_t and S_{t-1} used for the “natural estimator”: $\Delta t_1 = 1$ min, $\Delta t_2 = 10$ min, $\Delta t_3 = 1$ h, and $\Delta t_4 = 1$ day. The set of thresholds employed to investigate the directional-change approach can also be arbitrarily chosen. However, we selected them with the intent to compare the results of both experiments. For this reason, we used the number of returns in the data sample corresponding to each time interval Δt_k as the target for the number of directional changes registered in the same data set. That is, the collection of four thresholds δ_k was selected in such a way that in the given time series the number of directional changes is approximately equal to the number of time intervals n_k of the length Δt_k . We utilised one of the scaling properties described in Glattfelder et al. (2011) to find the precise thresholds size. The scaling property has the name “time of total-move” scaling law (law 10 in the article). The total-move is composed as the sum of the directional-change (DC) and overshoot (OS) parts. The law connects the size of the threshold δ with the waiting time $T_{TM}(\delta)$ between two consecutive intrinsic events:

$$T_{TM}(\delta) = \left(\frac{\delta}{C_{t,TM}} \right)^{E_{t,TM}}, \tag{21}$$

where $C_{t, TM}$ and $E_{t, TM}$ are the scaling coefficients. Equation (21) can be used to express the threshold δ in terms of the waiting time T_{TM} :

$$\delta(T_{TM}) = T_{TM}^{1/E_{t, TM}} C_{t, TM}. \tag{22}$$

The currency average scaling parameters $E_{t, TM}$ and $C_{t, TM}$ computed in [Glattfelder et al. \(2011\)](#) are 2.02 and 1.65×10^{-3} , correspondingly. Putting these coefficients into Equation (21), one can calculate that thresholds reciprocal to the selected time intervals $\Delta t_1, \dots, \Delta t_4$ are: $\delta(\Delta t_1) = 0.013\%$, $\delta(\Delta t_2) = 0.039\%$, $\delta(\Delta t_3) = 0.095\%$, and $\delta(\Delta t_4) = 0.458\%$. It is worth mentioning that applied scaling parameters are relevant only to the FX market which was the object of the research in [Glattfelder et al. \(2011\)](#). To the extent of our knowledge, parameters specific to Bitcoin prices, as well as to the S&P500 index, were not mentioned in the scientific literature before. Therefore, as the first step, we obtained the parameters by studying the “time of total-move” scaling law of historical Bitcoin, and SPX500 returns. The log-log plot of waiting times $T_{TM}(\delta)$ versus the directional-change threshold size δ is provided in Figure 4. The red line marks BTC/USD scaling law and is shown together with black, yellow, and green lines computed for EUR/USD, SPX500, and Geometrical Brownian Motion (GBM) correspondingly. Settings of the latter are chosen to mimic returns typical for the FX market.

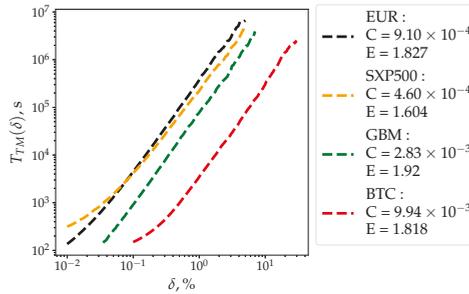


Figure 4. Time of total-move scaling laws computed for BTC/USD, EUR/USD, SPX500, and Geometrical Brownian Motion (GBM). GBM’s parameters are $S_0 = 1.3367$, $\mu = 0$, $\sigma = 20\%$, $T = 1$ year, and 10 million ticks in total. Scaling parameters C and E correspond to the coefficients of Equation (21).

Total-move scaling law parameters, obtained in the experiment, exhibit distinct resemblance of the stylised properties of the traditional FX and SPX500, as well as the emerging Bitcoin markets. Scaling factors $E_{t, TM}$ of EUR/USD, BTC/USD, and SPX500 are 1.827, 1.818, and 1.604, correspondingly. The coefficient specific for the BTC/USD pair is approximately 0.5% smaller than the one of EUR/USD. The coefficient of the SPX500 index is, in turn, is substantially smaller: by 9.9%. The same scaling factor of the GBM is the biggest among others: 1.920 ($\approx 5.6\%$ difference with EUR/USD). The parameter is noticeably distant from the parameters of the analysed exchange rates. We account the divergence to the non-normal distribution of real returns at ultra-short timescales (fat tails). The fat tails effect is pronounced in Figure 4 as the upward bend of the curves towards the beginning of the X-axis. The bends are read as the longer time needed for a total-move to unwrap than it is predicted by the linear part of the plot in the range of higher thresholds values. Linear regressions, built in the range of straight parts of the curves, are characterised by the scaling coefficients $E_{t, TM}$, which are close to the ones observed in GBM. The observed evidence is an additional confirmation of the “Aggregational Gaussianity” stylised fact¹³ typical for high-frequency markets ([Cont 2001a](#)). Scaling parameters $C_{t, TM}$ of EUR/USD, BTC/USD, and SPX500 are 9.07×10^{-4} , 9.94×10^{-3} ,

¹³ The evidence that the distribution of returns approaches the normal one measured over longer timescales.

and 4.60×10^{-4} , correspondingly. These values are significantly different due to the unlike scale of the corresponding volatility. This volatility dependent scaling parameter is not critical for the current analysis and will be discussed in the future research works.

The goal of the experiment is to compare the volatility computed using the “traditional” approach to the volatility based on the directional-change intrinsic time concept. Scaling law parameters $E_{t, TM}$ and $C_{t, TM}$ of historical BTC/USD returns were used to find the size of the directional-change thresholds, which would result in the average number of registered intrinsic events in the entire data-sample equal to the number of evenly spaced periods n_k . Expressing the parameter δ_k from the Equation (21) we find that for BTC/USD the thresholds are: $\delta(\Delta t_1) = 0.09\%$, $\delta(\Delta t_2) = 0.33\%$, $\delta(\Delta t_3) = 0.89\%$, $\delta(\Delta t_4) = 5.13\%$. The values are about ten times bigger than the ones related to the FX market (mentioned above) because of the proportionally larger realised volatility.

The same procedure, described in the previous paragraph, was performed in order to find the corresponding thresholds for the SPX500 time series. The obtained values: $\delta(\Delta t_1) = 0.006\%$, $\delta(\Delta t_2) = 0.025\%$, $\delta(\Delta t_3) = 0.076\%$, $\delta(\Delta t_4) = 0.55\%$.

The set of selected time intervals $\Delta t_{k=\{1, \dots, 4\}}$ and the complementary thresholds $\delta_{\Delta t_1, \dots, \delta_{\Delta t_4}}$ specific for each considered market were used to calculate realised and instantaneous volatility by traditional and the novel approach. We present in Table 2: average value of the realised volatility $\langle \sigma_{trad} \rangle$ computed as the sum of all four measurements ($k = \{1, 2, 3, 4\}$) divided by the number of experiments; its standard deviation σ_{trad}^- ; average value of the instantaneous volatility computed by the novel approach $\langle \sigma_{DC} \rangle$; the corresponding standard deviation σ_{DC}^- ; ratios of both measures $\langle \sigma_{trad} \rangle / \langle \sigma_{DC} \rangle$ and $\sigma_{trad}^- / \sigma_{DC}^-$. The last column of the table demonstrates the difference in the stability of results obtained by two measures.

The size difference of the realised and the instantaneous volatility is significant and is pronounced across all tested exchange rates (column $\langle \sigma_{trad} \rangle / \langle \sigma_{DC} \rangle$). The realised volatility computed in the “natural” way persistently exceeds the instantaneous volatility discovered via the novel approach. Only the two types of Bitcoin’s volatility appear to be 5% different whenever the divergence grows up to 99% in the case of SPX500. The striking difference is partially explained by the various discreteness of the employed data (which will be elaborated in the next section), and partially by the phenomenological properties of the selected markets (more on it in Section 7.4). This phenomenon is captivating especially taking into account that Bitcoin is particularly famous due to its oversized price activity. Its activity is clearly pronounced as the large standard deviation of the instantaneous volatility of BTC/USD pair (column σ_{DC}^-). Three FX exchange rates, having noticeably smaller realised volatility, are characterised by the wider range of the standard deviation values (column σ_{trad}^-). The ratio $\sigma_{trad}^- / \sigma_{DC}^-$ reaches the 0.02 level computed for EUR/USD. In other words, the standard deviation of the EUR/USD instantaneous volatility is 50 times bigger than the realised volatility value.

Table 2. Volatility of the considered time series computed using the “traditional” (Equation (20)) and the directional-change (Equation (19)) approaches. Provided values $\langle \sigma_{trad} \rangle$ and $\langle \sigma_{DC} \rangle$ are the average of four measurements performed with specific parameters: in the “traditional” case time intervals between observations S_n and S_{n-1} are $\Delta t_1 = 1$ min, $\Delta t_2 = 10$ min, $\Delta t_3 = 1$ h, and $\Delta t_4 = 1$ day. In the case of the directional-change intrinsic time approach, the thresholds δ are $\delta(\Delta t_1) = 0.013\%$, $\delta(\Delta t_2) = 0.039\%$, $\delta(\Delta t_3) = 0.095\%$, $\delta(\Delta t_4) = 0.458\%$ (FX prices), $\delta(\Delta t_1) = 0.09\%$, $\delta(\Delta t_2) = 0.33\%$, $\delta(\Delta t_3) = 0.89\%$, $\delta(\Delta t_4) = 5.13\%$ (BTC prices), and $\delta(\Delta t_1) = 0.006\%$, $\delta(\Delta t_2) = 0.025\%$, $\delta(\Delta t_3) = 0.075\%$, $\delta(\Delta t_4) = 0.545\%$ (SPX500).

Name	$\langle \sigma_{trad} \rangle, \%$	σ_{trad}^-	$\langle \sigma_{DC} \rangle, \%$	σ_{DC}^-	$\langle \sigma_{trad} \rangle / \langle \sigma_{DC} \rangle$	$\sigma_{trad}^- / \sigma_{DC}^-$
EUR/USD	9.72	0.03	7.53	1.38	1.29	0.02
EUR/JPY	11.93	0.12	8.55	2.07	1.40	0.06
EUR/GBP	8.04	0.23	5.81	1.43	1.38	0.16
BTC/USD	84.76	8.67	80.87	22.21	1.05	0.39
SPX500	13.19	0.67	6.63	3.24	1.99	0.21

7.3. Discrete Price Effect

The instantaneous volatility standard deviation computed for four different directional-change thresholds has an extremely high value (column σ_{DC}^- in Table 2). This indicates that in contrast to the realised volatility capable of the scaling, together with the time interval Δt , the instantaneous one does not scale together with the threshold size δ .

The price discontinuity typical for all real markets is the cause of the high standard deviation of the instantaneous volatility computed by the directional-change approach. Conventional exchange architecture restricts the price quotations to be a multiple of some constant, for example, 0.001 of a USD. This discreteness caused substantial debates in the scientific literature with regards to the accuracy of the "natural" estimators and on the extent to which they overestimate the actual volatility of the studied process (French and Roll 1986; Gottlieb and Kalay 1985). Equation (19) connects the number of directional changes observed per period of time and the instantaneous volatility and is built on the assumption of the continuous Brownian process. It has no adjustment factors to the discreteness of the analysed data. In reality, the directional-change intrinsic time does not precisely tick at the level where the size of the return is equal to the size of the threshold δ . Instead, in most cases, a new directional-change event becomes registered when the price has already jumped over the expected level. This discreteness effect becomes more pronounced when the size of the elementary price move (tick) is relatively big. That is, two factors contribute to the size of the instantaneous volatility of the discrete data-sample computed by the novel approach: the scale of the selected threshold δ and the tick size in the given sample (discreteness). Further, we provide results of a set of experiments where the impact of the price discreteness and the threshold size δ on the computed instantaneous volatility is observed.

Three time series were generated by GBM with the various density of ticks per period of time. Variation of the number of price changes in the simulation is equivalent to changing the simulated tick size having fixed a one-year time interval and volatility of the generated process fixed to be 15%. Forty gradually increasing directional-change thresholds were applied to all three GBMs. The thresholds range from 0.01% to 0.29%. We provide the plot of the computed instantaneous volatility of simulated time series in Figure 5. Two particular properties can be noticed. First, one brings the generated time series closer to the continuous process by making the size of a tick smaller (increasing the number of price changes per period in the sample). In this case, the values of the estimated instantaneous volatility σ_{DC} become closer to the volatility σ embedded in the model. Second, bigger thresholds are less sensitive to the discreteness of the given set of prices. The slippage effect of the price jump over the expected intrinsic time level becomes less pronounced, and the obtained result also approaches the value σ when the tick size represents a small fraction of the directional-change thresholds. A more comprehensive analysis should be performed in further research works to bridge the gap between the realised and instantaneous volatilities.

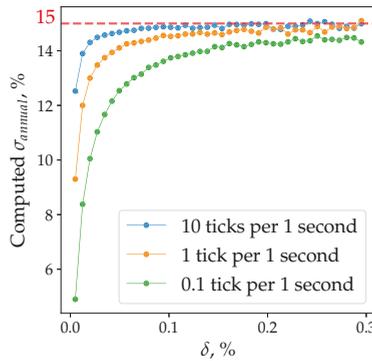


Figure 5. Instantaneous volatility of three time series generated by GBM with various tick frequencies and fixed volatility (15%). The volatility is computed by the directional-change approach (Equation (19)). Sizes of the directional change thresholds, used to calculate the volatility values, are put on the X-axis. Red dashed line marks the 15% level.

7.4. Volatility Seasonality

Dacorogna et al. (1993) presented a weekly seasonality pattern of price activity in the FX market. The authors’ analysis is based on the assumption that worldwide trading happens at strictly separated time zones with several dominated cores and operates within specific trading hours. Such a physical distribution of traders is embodied in geographical components of the market activity and eventually becomes pronounced as the weekly volatility seasonality. We do not build a similar assumption in our work. Instead, the collection of observed historical returns is treated as the only source of information available for the analysis. Further, we discover and describe the seasonality pattern of instantaneous volatility typical for FX exchange rates, Bitcoin prices, and S&P500 index.

We divide a whole week into a set of 10-min time intervals (bins). There are 1008 equally spaced points located at the fixed distance from the beginning of each week. This is a significantly larger number than the one used in the work Dacorogna et al. (1993) (168 points spaced by one hour intervals). We can afford this decreased granularity thanks to the more detailed historical time series employed for the experiment: instead of 12 million ticks for 26 exchange rates, we have on average 100 million ticks for each of the FX pairs. For each bin, the average number of directional changes will be computed.

The following series of steps allowed to construct the seasonality pattern. First, we run all historical tick-by-tick prices through the directional-change algorithm with the specified threshold δ . As soon as a new intrinsic event becomes registered, we check within which out of 1008 bins it happened. We add +1 to the number of directional changes corresponding to that time interval. When there are no prices left in the historical time series, we find the average number of intrinsic events per each bin. Equation (19) is then applied to compute the corresponding instantaneous volatility. Considering the five-year-long historical data, the obtained average is based on nearly 250 observations. Calculated instantaneous volatility values should be later normalised by the number of years and the length of a bin to get the annualised volatility specific for each bin of the week.

We select the threshold $\delta = 0.01\%$ for the first experiment with FX exchange rates. The average number of directional changes in a week registered by a threshold of this size is approximately equal to the number of 10-min long bins in it (1008). The reconstructed instantaneous volatility seasonality pattern of the FX pairs is shown in Figure 6. The pattern is notably stable across all tested exchange rates and is similar to the one demonstrated in Dacorogna et al. (1993).

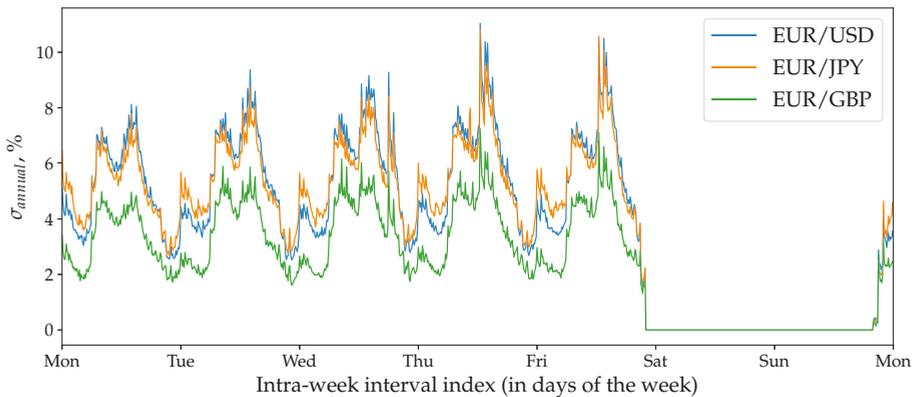


Figure 6. Instantaneous volatility seasonality of three Forex (FX) exchange rates computed using the directional-change approach (Equation (19)). Applied directional-change threshold $\delta = 0.01\%$. The whole week is divided by equally spaced time intervals $T = 10$ min (1008 bins in total).

We provide results of the same experiment where the “traditional” volatility estimator (Equation (20)) was employed to reveal the seasonality patterns of the FX exchange rates in Figure 7 and of BTC/USD in Figure 8. In contrast to the volatility seasonality pattern computed using the directional-change approach (Figure 6), the “traditional” pattern is less affected by the frequency of ticks per period of time specific for each studied time series. The difference between the average realised volatility across a week of the most active pair (EUR/JPY) and the least active (EUR/GBP)¹⁴ is equal to 46%. The same difference of the instantaneous seasonality (Figure 6) is 10% bigger and equal to 56%. The “traditional” estimator of the realised volatility seasonality demonstrates more rapid changes in the consecutive bins values. Local maximums at the beginning and the end of a day are considerably abrupt. The reason for this is that the directional-change intrinsic time captures the part of the volatility of the underlying process free of the noise component by ignoring the overshoot part of each trend move. The exact form and scale of the noise component and its connection to the overshoot section of the directional-change intrinsic time is a topic for future research work.

Assets traded in the crypto market have several specific properties which make them noticeably different compared to the traditional financial instruments such as FX exchange rates. Among the characteristics are: open trading within weekends and holidays; the absence of isolated physical trading centres where working hours are fixed; still low acceptance of the emerging market among professional traders. The outlined differences are endorsed by the history of technologies employed in the traditional FX and the emerging crypto worlds. The first one has originated in times when the trading happened in person and the settlement assumed the actual physical assets delivery. The trading organically evolved over time and became digital thanks to the internet expansion. Nevertheless, old properties, such as the governmental and the middle-man controls, have never been removed from the list of the accompanying FX markets design features. Bitcoin, in turn, has been designed as the alternative of the traditional financial system. It benefits from the blockchain technology by endorsing the principles of equality, openness, and accountability. We studied the historical prices of Bitcoin to investigate whether these specialities have any considerable impact on the BTC/USD instantaneous volatility seasonality pattern.

¹⁴ According to the Table 2.

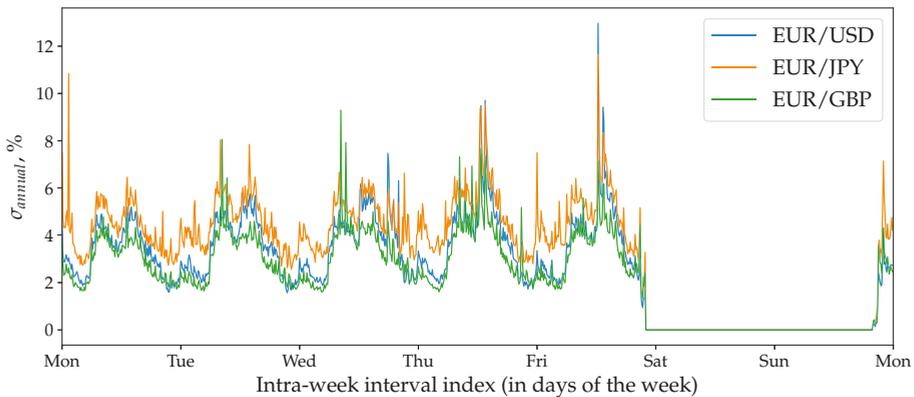


Figure 7. Realised volatility seasonality patterns of three FX exchange rates computed using the traditional approach (Equation (20)). Time intervals of 1-min have been used to calculate returns. The size of each bin is 10 min, 1008 bins in total.

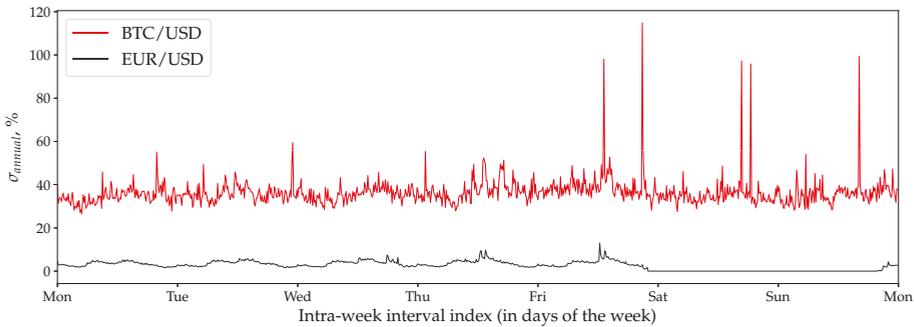


Figure 8. Realised volatility seasonality patterns of BTC/USD and EUR/USD exchange rates computed using the traditional approach (Equation (20)). Time intervals of 1-min have been used to calculate returns. The size of each bin is 10 min, there are 1008 bins in total.

We apply the same threshold size $\delta = 0.01\%$ used in the FX experiment to compare the seasonality patterns of Bitcoin and EUR. The obtained seasonality pattern put on top of the EUR/USD seasonality is presented in Figure 9. As can be seen from Figure 9, the periodical shape of Bitcoin’s curve is much less pronounced in contrast to EUR/USD. Its standard deviation computed within a week is 0.5%. It is roughly four times smaller than the standard deviation of the EUR/USD pattern (equal to 1.9%). Surprisingly, the intra-day maximums and minimums of Bitcoin seasonality do not precisely coincide with those observed in the traditional market. They are shifted towards the time intervals where European and American markets contribute the most to the geographical pattern (as disclosed in [Dacorogna et al. \(1993\)](#)). This observation confirms the one provided in [Eross et al. \(2017\)](#). That is particularly interesting since Asian markets are known for their substantial contribution to the cryptocurrency trading volumes. The fact that China has ruled that financial institutions cannot handle any Bitcoin transactions could be the reason of the observed phenomenon ([Ponsford 2015](#)). Instantaneous volatility over weekends is slightly lower than within the middle part of the week and is practically equal to Monday’s activity. We attribute the observed facts to the mentioned above non-traditional characteristic of the cryptocurrency market.

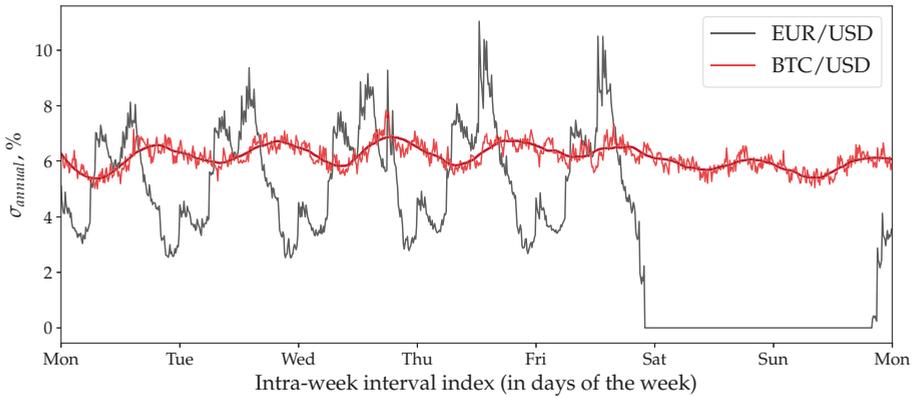


Figure 9. Instantaneous volatility seasonality of BTC/USD compared to the seasonality pattern of EUR/USD computed using the directional-change approach (Equation (19)). The dark-red curve approximates the Bitcoin seasonality pattern using the Savitzky–Golay filter (number of points in the window is 101, the order of the polynomial is 2). The directional-change threshold $\delta = 0.01\%$ was used in both experiments. Each discrete time interval (bin) is $T = 10$ min. There are 1008 bins in total.

As it was shown before, the instantaneous volatility computed by the novel approach directly depends on the size of the selected directional-change threshold δ (Figure 5). To examine the threshold size impact on the seasonality pattern of the real data, we arbitrarily selected the following set of values: $\delta = \{0.01\%, 0.04\%, 0.10\%\}$. The same algorithm described above was applied to reconstruct the volatility seasonality pattern for the FX pair EUR/USD (Figure 10) and SPX500 (Figure 11). The seasonality patterns shift toward higher volatility values when the size of the threshold is bigger. The observation is in line with the results of the experiments on GBM (Figure 5). Average values of EUR/USD seasonality curves computed with thresholds δ equal to 0.10% and 0.04% are correspondingly 1.71 and 1.57 times higher than the values computed with $\delta = 0.01\%$. The dependence of the seasonality smoothness on the size of the directional-change threshold become vividly pronounced: the seasonality curve constructed with the smallest threshold in the set is much sleeker (less wander) than the rest of the curves. This phenomenon should urge researchers and practitioners to select directional-change thresholds according to their needs very carefully while employing the directional-change technique.

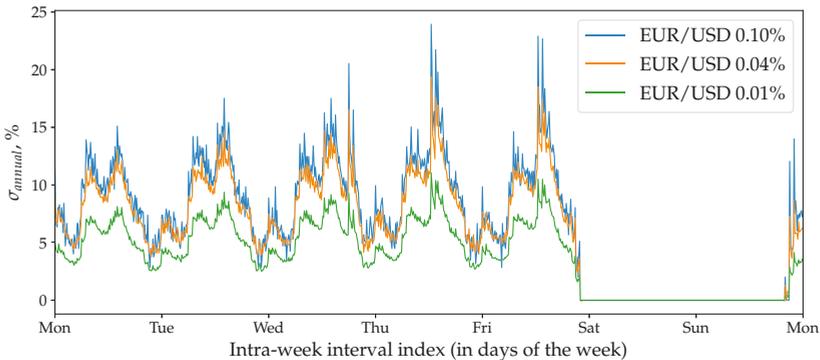


Figure 10. Volatility seasonality of EUR/USD computed using the novel approach (Equation (19)) and three different thresholds: $\delta = \{0.01\%, 0.04\%, 0.10\%\}$. The size of a bin is 10 min, there are 1008 bins in a week.

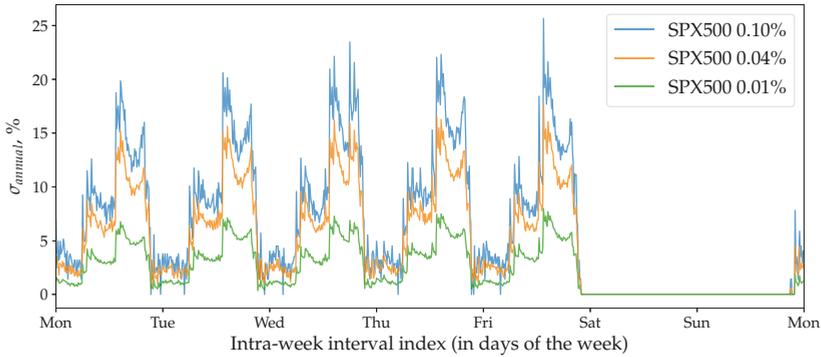


Figure 11. Volatility seasonality of SPX500 computed using the novel approach (Equation (19)) and three different thresholds: $\delta = \{0.01\%, 0.04\%, 0.10\%\}$. The size of a bin is 10 min, there are 1008 bins in a week.

According to Table 2, realised volatility of Bitcoin returns computed in the “traditional” way is about nine and six times bigger than the analogous volatility of the FX and SPX500 exchange rates (column $\langle \sigma_{trad} \rangle$). Besides, the retrieved sample of historical BTC/USD prices has 1.2 million ticks per year, which is 16.7 times smaller than the number of ticks per year in the EUR/USD case (about 20 million). As a result, the choice of the directional-change threshold δ has a much more significant effect on the average BTC/USD instantaneous volatility. We demonstrate results of four experiments where different threshold sizes were employed to reveal the seasonality patterns in Figure 12. The same $\delta = 0.01\%$ was used as the reference for the set of all arbitrary selected thresholds: $\delta = \{0.01\%, 0.03\%, 0.10\%, 0.20\%\}$. As it can be seen from Figure 12, the increase in the size of δ causes the corresponding increase in the volatility level around which the seasonality patterns oscillate. The levels of the seasonality distribution for $\delta = \{0.03\%, 0.10\%, 0.20\%\}$ are 1.5, 4.0, and 11.1 times bigger than the value corresponding to the smallest threshold $\delta = 0.01\%$. The biggest $\delta = 0.20\%$ lifts the value up to the level of $\sigma_{annual} = 68.5\%$ (which is still smaller than the realised volatility presented in Table 2 ($\sigma_{annual} = 84.76\%$)).

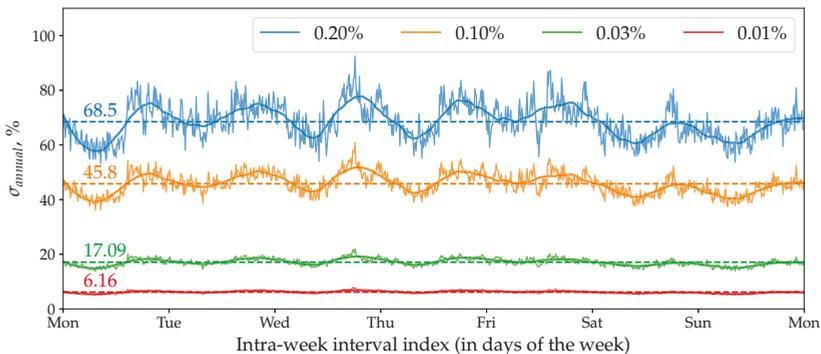


Figure 12. Instantaneous volatility seasonality of BTC/USD exchange rate computed using the directional-change approach (Equation (19)) and four different thresholds. Applied thresholds, from top to bottom: $\delta = \{0.20\%, 0.10\%, 0.03\%, 0.01\%\}$. The dark solid curves approximate the Bitcoin seasonality patterns using the Savitzky–Golay filter (number of points in the window is 101, order of the polynomial is 2). Bin size $T = 10$ min was chosen in all cases (1008 bins in a week). Dashed lines and the numbers over them represent the average level of each seasonality pattern across a week.

More information on the daily instantaneous and realised volatility seasonality ratio is provided in Figures A1–A4 and Table A1 in Appendix A.

7.5. Volatility Autocorrelation and Theta Time

The shape of the persistent instantaneous volatility seasonality patterns computed for the FX and SPX500 exchange rates changes with clear daily periodicity. This observation suggests that there should be a strong autocorrelation of the instantaneous volatility over time. The connection of the number of directional changes and the volatility value (Equation (19)) translates into the autocorrelation of the number of directional changes. We examined the autocorrelation function (ACF) of the number of directional changes observed within each bin of a week to check the assumption. The results of the experiment made for the FX exchange rates are provided in Figure 13. The same size of the directional-change threshold used to reveal the seasonality distribution $\delta = 0.01\%$ was employed. A remarkably stable pattern was found where daily and weekly seasonality is easily recognisable. The ACF function of FX exchange rates discovered in our work is highly similar to the results provided by Dacorogna et al. (1993). Nevertheless, there are clear differences between the FX autocorrelation patterns in our work and in the work of Dacorogna et al. (1993). The ACF of the number of directional changes computed through time lags defined in physical time does not cross the zero level for a much more extended period. It is consistently positive with lags even greater than several weeks. The curve representing the ACF of EUR/JPY has the smallest amplitude (smallest variability). In contrast, curves of EUR/USD together with EUR/GBP invariably follow the same pattern shifted up in the case of EUR/USD.

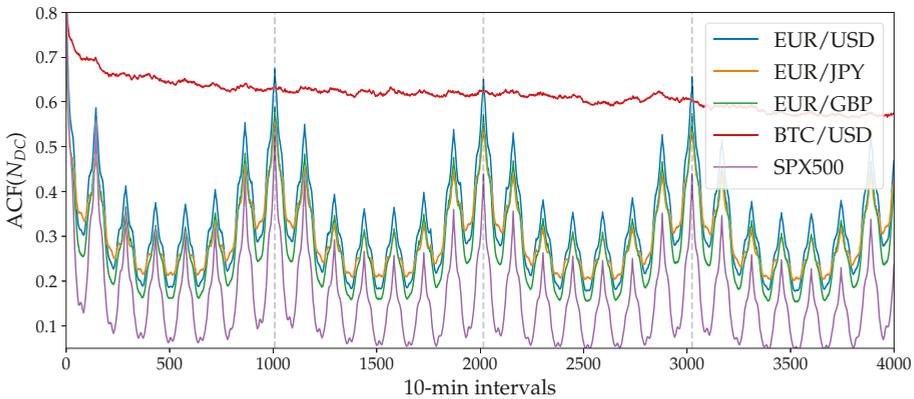


Figure 13. Autocorrelation function of the number of directional changes per 10-min long bins computed in physical time. Vertical dashed lines label weekly intervals. Applied threshold $\delta = 0.01\%$.

The SPX500 time series ACF has a similar shape to the shape of the FX market ACF. Two distinct properties can be noticed: the higher ACF amplitude and the faster decay. The exact level of decline of all exchange rates will be discussed a few paragraphs later.

The BTC/USD exchange rate seasonality pattern, characterised by much less pronounced instantaneous volatility, has also been tested in order to get the shape of the autocorrelation function. The results are presented in Figure 13. The amplitude of the ACF curves is the main difference in the values computed for the traditional FX and the emerging BTC/USD markets: the variability of the BTC/USD curve is 10 times lower than the variability of the EUR/USD one. We note that significantly bigger thresholds than the one used in the experiment ($\delta = 0.01\%$) have also been tested. All results confirmed that they reveal less accurate patterns due to the data insufficiently frequent for the statistical analysis.

A certain level of decline characterises the ACFs of all exchange rates as it can be seen from Figure 13. Large seasonal peaks of the autocorrelation functions drawn against of physical time do not allow to measure the level of decline precisely. A measure capable of converting the stochastic price evolution process to the stationary one should be applied to estimate the level of the downturn better. We minimise the seasonality pattern by employing the concept of theta time (Θ -time) proposed by Dacorogna et al. (1993). Θ -time is designed to eliminate the periodicity pattern by defining a set of non-equal time intervals within which the measure should be performed. The length of each Θ interval in physical time depends on the historical activity of the market. The theta time concept states that the average cumulative price activity (or volatility) between each consecutive couple of Θ steps is constant. Therefore, the distance between Θ timestamps, measured in physical time, is dictated by the shape of the volatility seasonality pattern. The periods of high price curve activity are equivalent to increasing the speed of physical time. The frequency of Θ stamps increases when the volatility rises too. In contrast, periods of low activity are identical to stretching the flow of the physical time, and the lower number of Θ intervals appears. As a result, active parts of the seasonality pattern, coinciding with the middle of the trading day, have the higher density of Θ timestamps per a unite of the physical time than the standstill sections overnights. Mathematically:

$$\Theta(t) = \int_{t_0}^t \sigma(t') dt', \tag{23}$$

where t_0 and t are the beginning and the end of the considered period of physical time and $\sigma(t')$ is the value of the instantaneous volatility corresponding to each moment of the interval. Equation (23) can be transformed into the sum of elements $\sigma_{\Delta t'}$ between the beginning and the end of the observed interval Δt_n in the case of a non-continuous seasonality pattern where the values are discretely defined in periods Δt (as in our experiment):

$$\Theta(t) = \sum_{\Delta t_0}^{\Delta t} \sigma_{\Delta t_n}. \tag{24}$$

It should be noted that the number of bins in a week is always constant in both physical and Θ times. This is achieved through the assumption that the integral (or the sum) of the weekly activity is the constant value.

The autocorrelation function of the number of directional changes computed in Θ -time is shown in Figure 14. Curves are approximated by the logarithmic function $y = A_{ACF} \log x + B_{ACF}$. The logarithmic coefficients A_{ACF} and B_{ACF} are presented on Figure 14 and in Table 3.

Major weekly fluctuations of the volatility seasonality pattern have been successfully eliminated for all three FX and BTC/USD exchange rates. Nevertheless, Θ -time does not completely remove the seasonality shape of ACF in the same way it happened in the work Dacorogna et al. (1993): noticeable peaks are still present in the final part of each business day. Moreover, the SPX500 curve is characterised by vividly pronounced daily seasonality pattern despite being run through the theta time algorithm. The phenomenon, observed in the original paper (Dacorogna et al. 1993), was explained by the non-optimal setup of the chosen model. The assumed same activity for all working days is indeed not fully correct (see Figure 6). However, we do not use any analytical expression postulating equal daily activity to describe the seasonality pattern. Instead, components $\sigma_{\Delta t'}$ of real empirically found volatility seasonality patterns depicted in Figures 6 and 9 were utilised to define the timestamps in Θ -time. Therefore, we eliminate the inefficiency connected to the assumption mentioned above. Thus, the alternative interpretation for the remained seasonality should be provided.

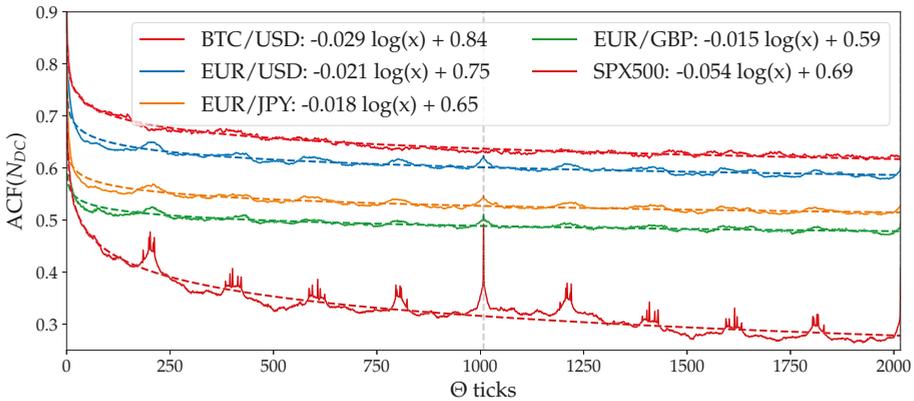


Figure 14. Autocorrelation function (ACF) of the number of directional changes per a bin in Θ -time. Vertical dashed line labels one week interval. There are 1008 bins in a week.

Table 3. Parameters of the logarithmic decay $y = A_{ACF} \log x + B_{ACF}$ used to fit the autocorrelation function (ACF) of the number of directional changes in Θ -time (Figure 14).

Name	A_{ACF}	B_{ACF}
BTC/USD	-0.029	0.84
EUR/USD	-0.021	0.75
EUR/JPY	-0.018	0.65
EUR/GBP	-0.015	0.59
SPX500	-0.054	0.69

We attribute the remaining fluctuations to the selected directional-change algorithm, which dissects the price curve into a collection of alternating trends. We also claim that the choice of the frequency of bins in a week used for the experiments affects the shape of the autocorrelation function in theta time. According to the directional-change algorithm (see Section 1), the dissection procedure has to be initialised only once and then it performs unsupervised. The evolution of the price curve dictates the sequence of intrinsic events. This fact leads to a certain dilemma: once registered, to which bin of a week should the intrinsic event be assigned? The following example illustrates the predicament. A couple of prices, at which two subsequent directional changes become registered, could belong to different bins. Let us say these are the intervals Δt_{n-1} and Δt_n . This means that the beginning of the price move that triggered the latest intrinsic event had started within Δt_{n-1} . But the end of this price trajectory finishes within the interval Δt_n . The crucial point is at what part of the Δt are the beginning and the end located. In the extreme case, the whole price trajectory before the directional change could be fully placed inside of the interval Δt_{n-1} . The latest tick that eventually triggered the new directional-change event can be at the very beginning of Δt_n . Should such an event be assigned to the bin Δt_{n-1} or to Δt_n ? The answer to this question is particularly important considering the effect the threshold size has on the seasonality patterns (Figures 10 and 11). The patterns constructed by using different thresholds have not only different average value over a week but also characterised by slightly shifted regions of local maximums and minimums (see, for example, the curves for $\delta = 0.01\%$ and $\delta = 0.10\%$).

A better way of associating locations of intrinsic events with bins of a week is another question related to the transition from the physical to intrinsic time and vice versa. This topic should be discussed in more details in further research works. Until then, the use of smaller thresholds and bigger time intervals is the strategy capable to impact the localisation problem positively.

8. Concluding Remarks

The language of traditionally considered drawdowns and drawups has been translated into the language of the directional-change intrinsic time. This translation made it possible to interpret the evolution of a price curve as a sequence of alternating trends of the given scale. The observed number of directional changes per period of time has been connected to the properties of the studied time series characterised by the instantaneous volatility σ_{DC} and the trend μ . The choice of directional-change thresholds δ_{up} and δ_{down} used to dissect the historical price curve is arbitrary but affects the results of the experiment. Bigger thresholds tend to register higher instantaneous volatility than the smaller ones. Equations (10), (11) and (16), connecting the observed number of directional changes to the properties of the studied process, have been validated by a Monte Carlo simulation. The simulation confirmed the robustness and accuracy of the obtained analytical expressions.

We extended the work of [Dacorogna et al. \(1993\)](#) by discovering the instantaneous volatility weekly seasonality pattern. One representative of the emerging cryptocurrency markets, Bitcoin, as well as of the traditional and widely accepted FX (EUR/USD, EUR/GBP, EUR/JPY) and stock (S&P500) markets, were considered. The connection of the number of directional-change intrinsic events to the instantaneous volatility has been employed to perform the computation. BTC/USD and SPX500 “total-move scaling laws” were computed for the first time to facilitate the seasonality discovery.

Similar patterns of the realised and the instantaneous volatility were obtained. Several noticeable differences between the results demonstrated in the work [Dacorogna et al. \(1993\)](#) and the ones presented in the current paper have been highlighted. First, the method based on the directional-change intrinsic time concept significantly simplifies the construction of the instantaneous volatility seasonality pattern operating with the information of the highest resolution (tick-by-tick prices). Second, the autocorrelation function of the number of directional changes computed in physical time stays positive for a notably long period of time. Third, the beginning of the volatility autocorrelation function computed in Θ -time can be approximated by the logarithmic function. The part of the autocorrelation plot after the lag bigger than one week declines linearly.

The difference between the instantaneous volatility seasonality patterns of the traditional FX and emerging Bitcoin markets demonstrates the currency digitisation and globalisation effect. The effect can be considered as the template for the future analogous markets to come.

The insights provided within this paper underline the relevance of the proposed directional-change framework as a valuable alternative to the traditional time-series analysis tools. The independence of directional-change intrinsic time on the frequency of price changes over a period of time makes it an effective tool for capturing periods of changing price activity. Results of the provided research can be used to extend the set of risk management tools constructed to evaluate the statistical properties of traditional and emerging financial markets.

Author Contributions: Conceptualisation and methodology, A.G., V.P. and R.O.; software and validation, V.P.; investigation, V.P. and A.G.; writing—original draft preparation, V.P.; writing—review and editing, V.P. and A.G.; and supervision, A.G. and R.O.

Funding: This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 675044.

Acknowledgments: We thank James Glattfelder, flov technologies, for all comments that greatly improved the manuscript. We are also grateful to the reviewers for all remarks.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Daily Seasonality

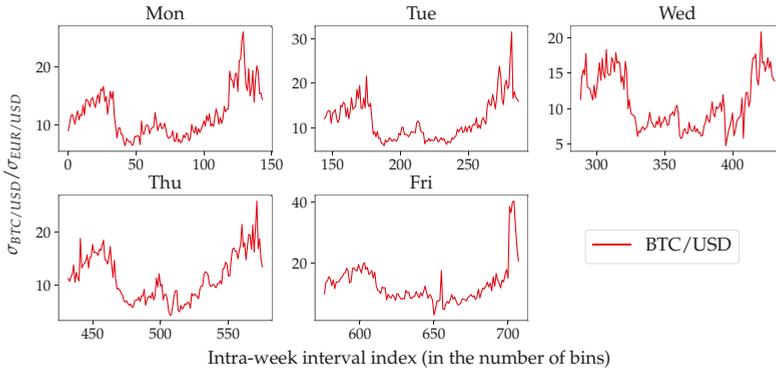


Figure A1. Daily realised volatility ratio of BTC/USD measured over 10-min time intervals dissecting the entire week into 1008 bins. The volatility is computed according to the “traditional” approach (Equation (20)).

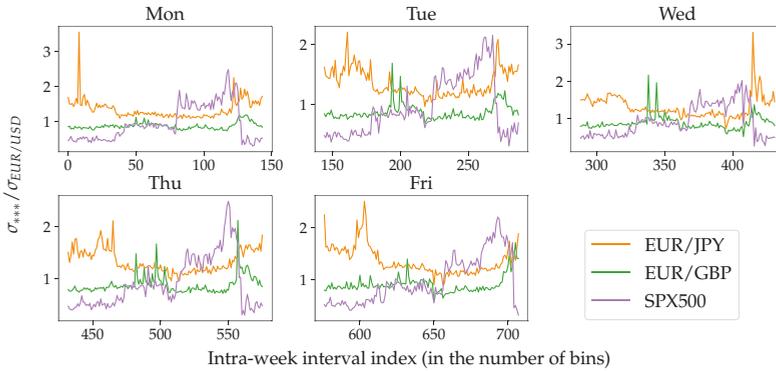


Figure A2. Daily realised volatility ratio of the given exchange rates (labelled by ***) and EUR/USD. The volatility is computed according to the “traditional” approach (Equation (20)).

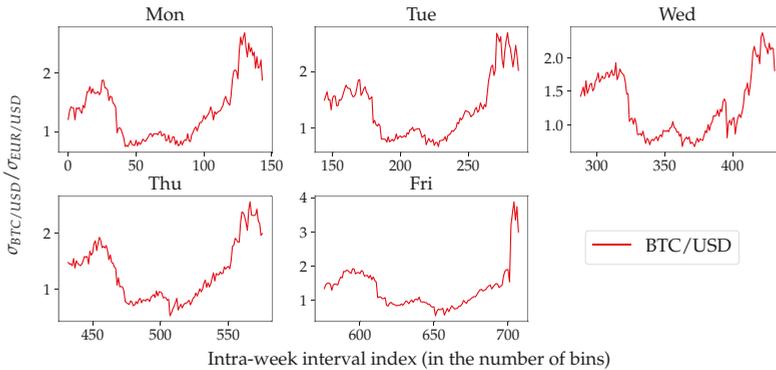


Figure A3. Daily instantaneous volatility ratio of BTC/USD measured over 10-min time intervals dissecting the entire week into 1008 bins. The volatility is computed according to the novel approach (Equation (19)).

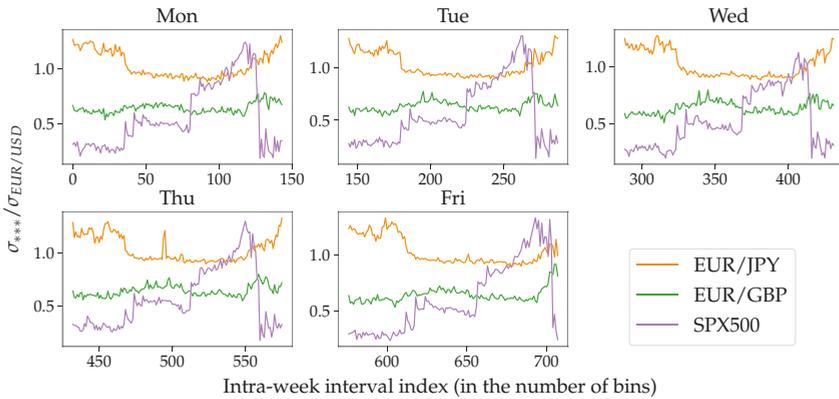


Figure A4. Daily instantaneous volatility ratio of the given exchange rates (labelled by ***) and EUR/USD. The volatility is computed according to the novel approach (Equation (19)).

Table A1. Daily instantaneous volatility ratio and the weekly standard deviation of two FX (EUR/JPY and EUR/GBP), one crypto (BTC/USD), and one stock (SPX500) exchange rates to EUR/USD. Columns $Ratio_*$ stands for the ratio of the average daily volatility of the corresponding exchange rate and one of EUR/USD. The average value computed over 10-min intervals. Columns std_* contain the standard deviation values of the volatility ratio over the set of 10-min intervals. The subscripts *trad* and *DC* label the measures made using the traditional volatility estimator (Equation (20)) and the novel approach (Equation (19)) correspondingly. The daily volatility ratios are graphically presented in Figures A2 and A3.

		$Ratio_{trad}$	std_{trad}	$Ratio_{DC}$	std_{DC}	$Ratio_{trad}/Ratio_{DC}$	std_{trad}/std_{DC}
Monday	EUR/JPY	1.35	0.27	1.03	0.12	1.31	2.25
	EUR/GBP	0.89	0.1	0.64	0.04	1.39	2.50
	BTC/USD	11.75	3.99	1.35	0.51	8.70	7.82
	SPX500	0.97	0.51	0.57	0.29	1.70	1.76
Tuesday	EUR/JPY	1.35	0.23	1.02	0.11	1.32	2.09
	EUR/GBP	0.88	0.13	0.63	0.05	1.40	2.60
	BTC/USD	11.7	4.53	1.35	0.52	8.67	8.71
	SPX500	0.95	0.47	0.57	0.3	1.67	1.57
Wednesday	EUR/JPY	1.31	0.3	1.02	0.12	1.28	2.50
	EUR/GBP	0.88	0.19	0.64	0.05	1.38	3.80
	BTC/USD	10.83	3.74	1.29	0.46	8.40	8.13
	SPX500	0.93	0.42	0.53	0.26	1.75	1.62
Thursday	EUR/JPY	1.34	0.22	1.03	0.12	1.30	1.83
	EUR/GBP	0.89	0.18	0.64	0.05	1.39	3.60
	BTC/USD	11.42	4.42	1.29	0.5	8.85	8.84
	SPX500	0.93	0.49	0.57	0.28	1.63	1.75
Friday	EUR/JPY	1.35	0.28	1.03	0.12	1.31	2.33
	EUR/GBP	0.9	0.17	0.64	0.06	1.41	2.83
	BTC/USD	11.81	4.93	1.3	0.54	9.08	9.13
	SPX500	0.98	0.46	0.61	0.31	1.61	1.48

References

A1, Yacine, and Robert Kimmel. 2007. Maximum likelihood estimation of stochastic volatility models. *Journal of Financial Economics* 83: 413–52.

Andersen, Torben G., and Jesper Lund. 1997. Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics* 77: 343–77. [CrossRef]

Asmussen, Søren, and Hansjörg Albrecher. 2010. *Ruin Probabilities*. Singapore: World Scientific Publishing Co Pte Ltd.

- Bandt, Christoph, and Faten Shiha. 2007. Order patterns in time series. *Journal of Time Series Analysis* 28: 646–65. [CrossRef]
- Barndorff-Nielsen, Ole E., and Neil Shephard. 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64: 253–80. [CrossRef]
- Begušić, Stjepan, Zvonko Kostanjčar, H. Eugene Stanley, and Boris Podobnik. 2018. Scaling properties of extreme price fluctuations in bitcoin markets. *Physica A: Statistical Mechanics and Its Applications* 510: 400–6. [CrossRef]
- Bessembinder, Hendrik. 1994. Bid-ask spreads in the interbank foreign exchange markets. *Journal of Financial Economics* 35: 317–48. [CrossRef]
- BIS. 2016. *Triennial Central Bank Survey: Foreign Exchange Turnover in April 2016*. Basel: BIS.
- Black, Fischer, and Myron Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81: 637–54. [CrossRef]
- Bollerslev, Tim, and Ian Domowitz. 1993. Trading patterns and prices in the interbank foreign exchange market. *The Journal of Finance* 48: 1421–43. [CrossRef]
- Bollerslev, Tim, and Michael Melvin. 1994. Bid—Ask spreads and volatility in the foreign exchange market: An empirical analysis. *Journal of International Economics* 36: 355–72. [CrossRef]
- Campbell, John Y., Stefano Giglio, Christopher Polk, and Robert Turley. 2018. An intertemporal capm with stochastic volatility. *Journal of Financial Economics* 128: 207–33. [CrossRef]
- Carr, Peter, Hongzhong Zhang, and Olympia Hadjilidiadis. 2011. Maximum drawdown insurance. *International Journal of Theoretical and Applied Finance* 14: 1195–30. [CrossRef]
- Chekhlov, Alexei, Stanislav Uryasev, and Michael Zabarankin. 2005. Drawdown measure in portfolio optimization. *International Journal of Theoretical and Applied Finance* 8: 13–58. [CrossRef]
- Cho, D Chinyung, and Edward W Frees. 1988. Estimating the volatility of discrete stock prices. *The Journal of Finance* 43: 451–66. [CrossRef]
- Cont, Rama. 2001a. *Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues*. Didcot: Taylor & Francis.
- Cont, Rama. 2011b. Statistical modeling of high-frequency financial data. *IEEE Signal Processing Magazine* 28: 16–25. [CrossRef]
- Cotter, John. 2005. Tail behaviour of the euro. *Applied Economics* 37: 827–40. [CrossRef]
- Dacorogna, Michael M., Ulrich A. Müller, Robert J. Nagler, Richard B. Olsen, and Olivier V. Pictet. 1993. A geographical model for the daily and weekly seasonal volatility in the foreign exchange market. *Journal of International Money and Finance* 12: 413–38. [CrossRef]
- Dacorogna, Michel M., Ulrich A. Müller, Olivier V. Pictet, and Casper G. De Vries. 2001. Extremal forex returns in extremely large data sets. *Extremes* 4: 105–27. [CrossRef]
- Dassios, Angelos, and Jia Wei Lim. 2018. An efficient algorithm for simulating the drawdown stopping time and the running maximum of a brownian motion. *Methodology and Computing in Applied Probability* 20: 189–204. [CrossRef]
- De Bondt, Werner F. M., and Richard H. Thaler. 1987. Further evidence on investor overreaction and stock market seasonality. *The Journal of Finance* 42: 557–81. [CrossRef]
- de Vries, Jon Erik, and Halvor Aarhus Aalborg. 2017. What Can Explain the Price, Volatility and Traded Volume of Bitcoin? Master's dissertation, University of Stavanger, Stavanger, Norway.
- Dyhrberg, Anne Haubo. 2016. Bitcoin, gold and the dollar—A garch volatility analysis. *Finance Research Letters* 16: 85–92. [CrossRef]
- Dyhrberg, Anne H., Sean Foley, and Jiri Svec. 2018. How investible is bitcoin? Analyzing the liquidity and transaction costs of bitcoin markets. *Economics Letters* 171: 140–3. [CrossRef]
- Eross, Andrea, Frank McGroarty, Andrew Urquhart, and Simon Wolfe. 2017. The Intraday Dynamics of Bitcoin. *SSRN Electronic Journal*. Available online: <https://dx.doi.org/10.2139/ssrn.3013699> (accessed on 12 September 2018).
- Fang, Lily, Chunmei Lin, and Yuping Shao. 2018. School holidays and stock market seasonality. *Financial Management* 47: 131–57. [CrossRef]
- French, Kenneth R., and Richard Roll. 1986. Stock return variances: The arrival of information and the reaction of traders. *Journal of Financial Economics* 17: 5–26. [CrossRef]
- Gençay, Ramazan, Michel Dacorogna, Ulrich A. Muller, Olivier Pictet, and Richard Olsen. 2001. *An Introduction to High-Frequency Finance*. Amsterdam: Elsevier.

- Glattfelder, James B., Alexandre Dupuis, and Richard B. Olsen. 2011. Patterns in high-frequency fx data: Discovery of 12 empirical scaling laws. *Quantitative Finance* 11: 599–614. [CrossRef]
- Golub, Anton, Gregor Chliamovitch, Alexandre Dupuis, and Bastien Chopard. 2014. Multi-Scale Representation of High Frequency Market Liquidity. Available online: <http://dx.doi.org/10.2139/ssrn.2393428> (accessed on 11 September 2018).
- Golub, Anton, James Glattfelder, and Richard B. Olsen. 2017. The Alpha Engine: Designing an Automated Trading Algorithm. Available online: <http://dx.doi.org/10.2139/ssrn.2951348> (accessed on 11 September 2018).
- Gottlieb, Gary, and Avner Kalay. 1985. Implications of the discreteness of observed stock prices. *The Journal of Finance* 40: 135–53. [CrossRef]
- Grossman, Sanford J., and Zhongquan Zhou. 1993. Optimal investment strategies for controlling drawdowns. *Mathematical Finance* 3: 241–76. [CrossRef]
- Guillaume, Dominique M., Michel M. Dacorogna, Rakhil R. Davé, Ulrich A. Müller, Richard B. Olsen, and Olivier V. Pictet. 1997. From the bird's eye to the microscope: A survey of new stylized facts of the intra-daily foreign exchange markets. *Finance and Stochastics* 1: 95–129. [CrossRef]
- Gultekin, Mustafa N., and N. Bulent Gultekin. 1983. Stock market seasonality: International evidence. *Journal of Financial Economics* 12: 469–81. [CrossRef]
- Hadjiliadis, Olympia, and Jan Večeř. 2006. Drawdowns preceding rallies in the brownian motion model. *Quantitative Finance* 6: 403–9. [CrossRef]
- Haferkorn, Martin, and Josué Manuel Quintana Diaz. 2014. Seasonality and interconnectivity within cryptocurrencies-an analysis on the basis of bitcoin, litecoin and namecoin. In *International Workshop on Enterprise Applications and Services in the Finance Industry*. Cham: Springer, pp. 106–20.
- Harris, Larry. 2003. *Trading and Exchanges: Market Microstructure for Practitioners*. Oxford: Oxford University Press.
- Hartmann, Philipp. 1999. Trading volumes and transaction costs in the foreign exchange market: Evidence from daily dollar-yen spot data. *Journal of Banking & Finance* 23: 801–24.
- Hasbrouck, Joel. 2018. High-frequency quoting: Short-term volatility in bids and offers. *Journal of Financial and Quantitative Analysis* 53: 613–41. [CrossRef]
- Ito, Takatoshi, and Yuko Hashimoto. 2006. Intraday seasonality in activities of the foreign exchange markets: Evidence from the electronic broking system. *Journal of the Japanese and International Economies* 20: 637–64. [CrossRef]
- Jondeau, Eric, and Michael Rockinger. 2003. Testing for differences in the tails of stock-market returns. *Journal of Empirical Finance* 10: 559–81. [CrossRef]
- Kaya, Orçun, Jan Schildbach, and Deutsche Bank Ag. 2016. High-frequency trading. *Reaching the Limits, Automated Trader Magazine* 41: 23–27.
- Keim, Donald B. 1983. Size-related anomalies and stock return seasonality: Further empirical evidence. *Journal of Financial Economics* 12: 13–32. [CrossRef]
- Kelly, John L., Jr. 2011. A new interpretation of information rate. In *The Kelly Capital Growth Investment Criterion: Theory and Practice*. Singapore: World Scientific, pp. 25–34.
- Koning, Nico, Daniel T. Cassidy, and Rachid Ouyed. 2018. Extended model of stock price behaviour. *Journal of Mathematical Finance* 8: 1. [CrossRef]
- Landriault, David, Bin Li, and Hongzhong Zhang. 2015. On the frequency of drawdowns for brownian motion processes. *Journal of Applied Probability* 52: 191–208. [CrossRef]
- Liu, Ruiqing, Zhichao Shao, Guodong Wei, and Wei Wang. 2017. Garch model with fat-tailed distributions and bitcoin exchange rate returns. *Journal of Accounting, Business and Finance Research* 1: 71–75. [CrossRef]
- Lundberg, Filip. 1909. Über die theorie der ruck-versicherung. In *Trans VI International Congress Actuaries*. pp. 877–955. Available online: <https://ci.nii.ac.jp/naid/10004541215/> (accessed on 10 December 2018).
- Mancino, Maria Elvira, Maria Cristina Recchioni, and Simona Sanfelici. 2017. Estimation of instantaneous volatility. In *Fourier-Malliavin Volatility Estimation*. Cham: Springer, pp. 31–47.
- Menkhoff, Lukas, Lucio Sarno, Maik Schmeling, and Andreas Schrimpf. 2012. Carry trades and global foreign exchange volatility. *The Journal of Finance* 67: 681–718. [CrossRef]
- Mijatović, Aleksandar, and Martijn R. Pistorius. 2012. On the drawdown of completely asymmetric lévy processes. *Stochastic Processes and their Applications* 122: 3812–36. [CrossRef]

- Müller, Ulrich A., Michel M. Dacorogna, Rakhal D. Davé, Richard B. Olsen, Olivier V. Pictet, and Jacob E. von Weizsäcker. 1997. Volatilities of different time resolutions—Analyzing the dynamics of market components. *Journal of Empirical Finance* 4: 213–39. [CrossRef]
- Müller, Ulrich A., Michel M. Dacorogna, Richard B. Olsen, Olivier V. Pictet, Matthias Schwarz, and Claude Morgengegg. 1990. Statistical study of foreign exchange rates, empirical evidence of a price change scaling law, and intraday analysis. *Journal of Banking & Finance* 14: 1189–208.
- Nakamoto, Satoshi. 2008. *Bitcoin: A Peer-to-Peer Electronic Cash System*. Available online: <http://bitcoin.org/bitcoin.pdf> (accessed on 23 October 2018).
- Osborne, Maury F. M. 1959. Brownian motion in the stock market. *Operations Research* 7: 145–73. [CrossRef]
- Ponsford, Matthew P. 2015. A comparative analysis of bitcoin and other decentralised virtual currencies: Legal regulation in the people’s republic of china, canada, and the united states. *HKJ Legal Studies* 9: 29.
- Pospasil, Libor, Jan Vecer, and Olympia Hadjiliadis. 2009. Formulas for stopped diffusion processes with stopping times based on drawdowns and drawups. *Stochastic Processes and their Applications* 119: 2563–78. [CrossRef]
- Rachev, Svetlozar T., Christian Menn, and Frank J. Fabozzi. 2005. *Fat-Tailed and Skewed Asset Return Distributions: Implications for Risk Management, Portfolio Selection, and Option Pricing*. Hoboken: John Wiley & Sons, vol. 139.
- Rolski, Tomasz, Hanspeter Schmidli, Volker Schmidt, and Jozef L. Teugels. 2009. *Stochastic Processes for Insurance and Finance*. Hoboken: John Wiley & Sons, vol. 505.
- Roy, Souvik, and P. Venkateswaran. 2014. Online payment system using steganography and visual cryptography. Paper presented at 2014 IEEE Students’ Conference on Electrical, Electronics and Computer Science, Bhopal, India, March 1–2. Lakes-Bhopal: IEEE, pp. 1–5.
- Rozeff, Michael S., and William R. Kinney, Jr. 1976. Capital market seasonality: The case of stock returns. *Journal of Financial Economics* 3: 379–402. [CrossRef]
- Sapuric, Svetlana, and Angelika Kokkinaki. 2014. Bitcoin is volatile! Isn’t that right? Paper presented at International Conference on Business Information Systems, Larnaca, Cyprus, May 22–23. Cham: Springer, pp. 255–65.
- Schuhmacher, Frank, and Martin Eling. 2011. Sufficient conditions for expected utility to imply drawdown-based performance rankings. *Journal of Banking & Finance* 35: 2311–18.
- Seif, Mostafa, Paul Docherty, and Abul Shamsuddin. 2017. Seasonal anomalies in advanced emerging stock markets. *The Quarterly Review of Economics and Finance* 66: 169–81. [CrossRef]
- Shintate, Takuya, and Lukáš Pichl. 2019. Trend prediction classification for high frequency bitcoin time series with deep learning. *Journal of Risk and Financial Management* 12: 17. [CrossRef]
- Taylor, Howard M. 1975. A stopped brownian motion formula. *The Annals of Probability* 3: 234–46. [CrossRef]
- Zarowin, Paul. 1990. Size, seasonality, and stock market overreaction. *Journal of Financial and Quantitative Analysis* 25: 113–25. [CrossRef]
- Zhang, Hongzhong. 2015. Occupation times, drawdowns, and drawups for one-dimensional regular diffusions. *Advances in Applied Probability* 47: 210–30. [CrossRef]
- Zhang, Hongzhong, and Olympia Hadjiliadis. 2012. Drawdowns and the speed of market crash. *Methodology and Computing in Applied Probability* 14: 739–52. [CrossRef]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Article

Statistical Arbitrage with Mean-Reverting Overnight Price Gaps on High-Frequency Data of the S&P 500

Johannes Stübinger * and Lucas Schneider

Department of Statistics and Econometrics, University of Erlangen-Nürnberg, Lange Gasse 20, 90403 Nürnberg, Germany; lucas.schneider@fau.de

* Correspondence: johannes.stuebinger@fau.de

Received: 27 February 2019; Accepted: 26 March 2019; Published: 1 April 2019

Abstract: This paper develops a fully-fledged statistical arbitrage strategy based on a mean-reverting jump–diffusion model and applies it to high-frequency data of the S&P 500 constituents from January 1998–December 2015. In particular, the established stock selection and trading framework identifies overnight price gaps based on an advanced jump test procedure and exploits temporary market anomalies during the first minutes of a trading day. The existence of the assumed mean-reverting property is confirmed by a preliminary analysis of the S&P 500 index; this characteristic is particularly significant 120 min after market opening. In the empirical back-testing study, the strategy delivers statistically- and economically-significant returns of 51.47 percent p.a. and an annualized Sharpe ratio of 2.38 after transaction costs. We benchmarked our trading algorithm against existing quantitative strategies from the same research area and found its performance superior in a multitude of risk-return characteristics. Finally, a deep dive analysis shows that our results are consistently profitable and robust against drawdowns, even in recent years.

Keywords: computational finance; asset pricing models; overnight price gaps; financial econometrics; mean-reversion; statistical arbitrage; high-frequency data; jump-diffusion model

1. Introduction

Statistical arbitrage is a market-neutral strategy developed by a quantitative group at Morgan Stanley in the mid-1980s (Pole 2011). Following Hogan et al. (2004), the self-financing strategy describes a long-term trading opportunity that exploits persistent capital market anomalies to draw positive expected profits with a Sharpe ratio that increases steadily over time. Arbitrage situations are identified with the aid of data-driven techniques ranging from plain vanilla approaches to state-of-the-art models. In the event of a temporary anomaly, an arbitrageur goes long in the undervalued stock and short in the overvalued stock (see Vidyamurthy (2004), Gatev et al. (2006)). If history repeats itself, prices converge to their long-term equilibrium and an investor makes a profit. Key contributions are provided by Vidyamurthy (2004), Gatev et al. (2006), Avellaneda and Lee (2010), Bertram (2010), Do and Faff (2012), and Chen et al. (2017).

The available literature divides statistical arbitrage into five sub-streams, including the time-series approach, which concentrates on mean-reverting price dynamics. Since financial data are exposed to more than one source of uncertainty, it is surprising that there exist only a few academic studies that use a jump-diffusion model (see Larsson et al. (2013), Göncü and Akyildirim (2016), Stübinger and Endres (2018), Endres and Stübinger (2019ab)). In addition to mean-reversion, volatility clusters, and drifts, this general and flexible stochastic model is able to capture jumps and fat tails. First, Larsson et al. (2013) used jump-diffusion models to formulate an optimal stopping theory. Göncü and Akyildirim (2016) presented a stochastic model for the daily trading of commodity pairs in which the noise-term is driven by a Lévy process. Stübinger and Endres (2018) introduce a holistic pair selection and trading

strategy based on a jump-diffusion model. Recently, [Endres and Stübinger \(2019ab\)](#) derived an optimal pairs trading framework based on a flexible Lévy-driven Ornstein–Uhlenbeck process and applied it to high-frequency data. All these studies deal with intraday price dynamics and are therefore not in a position to take into account the impact of overnight price changes, an apparent deficit as information is published in media platforms 24 h a day, seven days a week.

This paper enhances the existing research in several aspects. First, our manuscript contributes to the literature by developing a fully-fledged statistical arbitrage framework based on a jump–diffusion model, which is able to capture intraday and overnight high-frequency price dynamics. Specifically, we detect overnight price gaps based on the jump test of [Barndorff-Nielsen and Shephard \(2004\)](#) and [Andersen et al. \(2010\)](#) and exploit temporary market anomalies during the first minutes of a trading day. The existence of the assumed mean-reverting property is confirmed by a preliminary analysis on the S&P 500 index; this characteristic is particularly significant 120 min after market opening. Second, the value-add of the proposed trading framework is evaluated by benchmarking it against well-known quantitative strategies in the same research area. In particular, we consider the naive S&P 500 buy-and-hold strategy, fixed threshold strategy, general volatility strategy, as well as reverting volatility strategy. Third, we perform a large-scale empirical study on the sophisticated back-testing framework of high-frequency data of the S&P 500 constituents from January 1998–December 2015. Our jump-based strategy produces statistically- and economically-significant returns of 51.47 percent p.a. appropriate after transaction costs. The results outperform the benchmarks ranging from –6.56 percent for the fixed threshold strategy to 38.85 percent for the reverting volatility strategy; complexity pays off. Fourth, a deep-dive analysis shows that our results are consistently profitable and robust against drawdowns even in the last part of our sample period, which is noteworthy as almost all statistical arbitrage strategies have suffered from negative returns in recent years (see [Do and Faff \(2010\)](#), [Stübinger and Endres \(2018\)](#)). The results pose a major challenge to the semi-strong form of market efficiency.

The remainder of this research study is structured as follows. Section 2 provides the theoretical framework applied in this study. In Section 3, we discuss the event study of the S&P 500 index. After describing the empirical back-testing framework in Section 4, we analyze our results and present key findings in Section 5. Finally, Section 6 gives final remarks and an outlook on future work.

2. Methodology

This section provides the theoretical construct of our statistical arbitrage strategy. Therefore, Section 2.1 describes the Barndorff–Nielsen and Shephard jump test (BNS jump test), which helps us to recognize jumps in our time series. The identification of overnight gaps is presented in Section 2.2.

2.1. Barndorff–Nielsen and Shephard Jump Test

We follow the theoretical framework of [Barndorff-Nielsen and Shephard \(2004\)](#) to detect overnight gaps. First, let us denote low-frequency returns as:

$$y_i = y^*(i\hbar) + y^*((i - 1)\hbar), \quad i = 1, 2, \dots, \tag{1}$$

where $y^*(t)$ denotes the log price of an asset after time interval $\{t\}_{t \geq 0}$ and \hbar represents a fixed time period, e.g., trading days. These low-frequency returns can be split up into M equally-spaced high-frequency returns of the following form:

$$y_{j,i} = y^*((i - 1)\hbar + \hbar j M^{-1}) + y^*((i - 1)\hbar + \hbar(j - 1)M^{-1}), \quad j = 1, 2, \dots, M. \tag{2}$$

If i denotes the i^{th} day, the j^{th} intra- \hbar return is expressed as $y_{j,i}$. Therefore, the daily return can be written as:

$$y_i = \sum_{j=1}^M y_{j,i}. \tag{3}$$

The BNS jump test of [Barndorff-Nielsen and Shephard \(2004\)](#) underlies the assumptions that prices follow a semi-martingale to ensure the condition of no-arbitrage and are generated by a jump-diffusion process of the following form and properties:

$$y^*(t) = y^{(1)*}(t) + y^{(2)*}(t), \tag{4}$$

where $y^*(t)$ describes the log price and $y^{(1)*}(t)$ represents the stochastic volatility semi-martingale process:

$$y^* = \alpha^* + m^*, \tag{5}$$

with α^* describing the trend term with locally-finite variation paths, following a continuous mean process of the security. The stochastic volatility process is represented through m^* , which is a local martingale and defined as:

$$m^* = \int_0^t \sigma(u) dW(u), \tag{6}$$

where W describes the Wiener process. The spot volatility process $\sigma^2(t)$ is locally restricted away from zero and specified as càdlàg, meaning that the process is limited on the left side, while it is everywhere right continuous. Furthermore, $\sigma(t) > 0$, and the integrated variance (IV) process:

$$\sigma^{2*}(t) = \int_0^t \sigma^2(u) du \tag{7}$$

satisfies $\sigma^{2*}(t) < \infty, \forall t < \infty$. Moreover, $y^{(2)*}(t)$ defines the discontinuous jump component as:

$$y^{(2)*}(t) = \sum_{i=1}^{N(t)} c_i, \tag{8}$$

with N representing a finite counting process, so that $N(t) < \infty, \forall t > 0$ and c_i denoting nonzero random variables. Putting all together, the process can be written as:

$$y^*(t) = \alpha^* + \int_0^t \sigma(u) dW(u) + \sum_{i=1}^{N(t)} c_i \tag{9}$$

consisting of a stochastic volatility component that models continuous price motions and a jump term that accounts for sudden price shifts and discontinuous price changes. It is assumed that σ and α^* are independent of W . From an economic point of view, [Rombouts and Stentoft \(2011\)](#) showed that neglecting the non-Gaussian features of the data, prices are estimated with large errors.

To conduct the BNS jump test, three volatility metrics need to be specified: The quadratic variation (QV), realized variance (RV), and bipower variation (BPV). QV is defined as:

$$[y^*](t) = \sigma^{2*}(t) + \sum_{i=1}^{N(t)} c_i^2, \tag{10}$$

with $\sigma^{2*}(t)$ denoting the integrated variance, presenting the quadratic variation of the continuous part of the semi-martingale process, while $\sum_{i=1}^{N(t)} c_i^2$ determines the quadratic variation of the jump component (see [Andersen et al. \(2001\)](#), [Barndorff-Nielsen and Shephard \(2002\)](#), [Andersen et al. \(2003\)](#), [Barndorff-Nielsen and Shephard \(2004\)](#), [Barndorff-Nielsen and Shephard \(2006\)](#)). Hence, this volatility measurement takes into account the total variation of the underlying jump-diffusion process.

The realized variance:

$$[y_M^*]_i^2 = \sum_{j=1}^M y_{j,i}^2 \tag{11}$$

functions as a consistent estimator of QV , where M determines the number of intraday returns for day i . This volatility measure sums up all squared intraday returns for any considered period.

Andersen and Bollerslev (1998), Andersen et al. (2001), and Barndorff-Nielsen and Shephard (2002) showed that RV equals QV for large M , yielding to the equation:

$$\text{plim}_{M \rightarrow \infty} RV_t = QV_t = \sigma^{2*}(t) + \sum_{i=1}^{N(t)} c_i^2. \tag{12}$$

BPV was introduced by Barndorff-Nielsen and Shephard (2004) as:

$$\{y^*\}^{[r,s]}(t) = \text{plim}_{\delta \rightarrow 0} \delta^{1-(r+s)/2} \sum_{j=1}^{\lfloor t/\delta \rfloor - 1} |y_j|^r |y_{j+1}|^s, \quad r, s \geq 0, \tag{13}$$

where every $\{\delta\}_{\delta > 0}$ periods of time observations exist in interval t . BPV is a consistent estimator of IV under the assumption of a semi-martingale stochastic volatility process with a jump component described by Equation (4). Under those assumptions and for $r > 0$ and $s > 0$ applies:

$$\mu_r^{-1} \mu_s^{-1} \{y^*\}^{[r,s]}(t) = \begin{cases} \int_0^t \sigma^{r+s}(u) d(u), & \max(r, s) < 2, \\ x^*(t), & \max(r, s) = 2, \\ \infty, & \max(r, s) > 2, \end{cases} \tag{14}$$

where $x^*(t)$ is a stochastic process, and μ is defined as:

$$\mu_x = E|u|^x = 2^{x/2} \frac{\Gamma\left(\frac{1}{2}(x+1)\right)}{\Gamma\left(\frac{1}{2}\right)}, \tag{15}$$

with $x > 0$, u following a standard normal distribution, while Γ denotes the complete gamma function.

Barndorff-Nielsen and Shephard (2004) focused on the special case of $r = s = 1$ leading to the following equation:

$$\mu_1^{-2} \{y_M^*\}_i^{[1,1]} = \mu_1^{-2} \sum_{j=1}^{M-1} |y_{j,1}|^1 |y_{j+1,1}|^1 \xrightarrow{p} \int_{h(i-1)}^{hi} \sigma^2(u) du. \tag{16}$$

Hence, BPV is for $r = s = 1$ a consistent estimator of the integrated volatility for the i^{th} period. Based on this case, the variation of the jump term can be isolated by subtracting BPV from RV :

$$[y_M^*]_i^2 - \mu_1^{-2} \{y_M^*\}_i^{[1,1]} \xrightarrow{p} \sum_{j=N(h(i-1))+1}^{N(hi)} c_j^2. \tag{17}$$

By calculating the difference between RV and BPV , we can separate the jump contribution to the variation of the asset price from the QV . Therefore, the volatility can be decomposed into its continuous and discontinuous components.

To identify jumps, we use the basic principles of the non-parametric BNS jump test and apply the ratio z-statistic from Huang and Tauchen (2005). This test statistic is adjusted for market noise and provides useful properties such as an appropriate size and a reasonable power. The evidence from the

Monte Carlo simulation also suggests that this z-test is fairly accurate in detecting real jumps and not easily fooled by market micro structure noises. The ratio test statistic:

$$Z_t = \frac{RV_t - BPV_t}{RV_t} \sqrt{\left(\left(\frac{\pi}{2}\right)^2 + \pi - 5\right) \frac{1}{M} \max\left(1, \frac{TP_t}{BV_t^2}\right)} \xrightarrow{d} N(0, 1) \text{ as } M \rightarrow \infty \tag{18}$$

is asymptotic standard normally distributed under the null hypothesis of no jumps. Following [Huang and Tauchen \(2005\)](#), the tripower quarticity statistic is calculated by the following equation:

$$TP_t = M\mu_{4/3}^{-3} \left(\frac{M}{M-2}\right) \sum_{j=3}^M |r_{t,j}|^{4/3} |r_{t,j-1}|^{4/3} |r_{t,j-2}|^{4/3} \rightarrow \int_0^t \sigma^4(u) du. \tag{19}$$

To determine if at least one jump occurred in an asset, a right-sided hypothesis test with the null hypothesis of no jumps was conducted. A commonly-used level of significance is 0.1 percent (see [Barndorff-Nielsen and Shephard \(2006\)](#), [Evans \(2011\)](#), [Frömmel et al. \(2015\)](#)). If the null hypothesis was rejected, at least one jump emerged in the underlying security during the considered period.

2.2. Jump Detection Scheme

The timing of jumps has an essential meaning for examining anomalous behavior around jumps. To identify overnight gaps via jump tests, the precise time must be known. For this purpose, we rely on the jump detection scheme introduced by [Andersen et al. \(2010\)](#). This jump identification procedure is designed on the premise that jumps are rare events. If it is assumed that t equals one day and at most one jump can emerge during the corresponding period, the only intraday jump can be determined with:

$$RV_t - BPV_t \xrightarrow{p} c_t^2, \tag{20}$$

where c_t^2 represents the jump variation in period t . The intuitive idea is that the jump must be incorporated in the highest absolute return on that specific day. Hence, the timing of the jump can be determined by seeking the highest absolute return of the period. Furthermore, the precise jump size can be calculated in the following way:

$$\tilde{c}_t = \text{sgn}\left(\left\{r_{t,c} : |r_{t,c}| = \max_{j \in \{1, \dots, M\}} |r_{t,j}|\right\}\right) c_t^2, \tag{21}$$

where $r_{t,c}$ denotes the intraday return that contains the jump contribution, while $\text{sgn}(\cdot)$ is equal to 1 or -1 , depending on the sign of the argument.

3. Event Study of the S&P 500 Index

This section uses the outlined methodology of Section 2 to identify and analyze overnight price gaps in the S&P 500 index. Following the approaches of [Fung et al. \(2000\)](#) and [Grant et al. \(2005\)](#), we conducted the following four steps.

At first, the data were filtered according to the event of interest, the presence of overnight gaps. To identify overnight gaps, we conducted daily the BNS jump tests, as introduced in Section 2.1. For the test, we used high-frequency intraday returns of the previous day and the overnight return and a significance level of 0.1 percent. The timing of jumps was determined by the jump detection procedure of [Andersen et al. \(2010\)](#) (see Section 2.2). If the timing of the jump corresponded with the overnight return, the day was marked as an event day and included in our study.

Second, for every event day, the cumulative return of the S&P 500 index at minute t after the market opening was computed by:

$$CR_{i,t} = \frac{P_{i,t}}{P_{i,0}} - 1, \tag{22}$$

where $P_{i,t}$ denotes the index price on event day i at minute t after the beginning of the trading day. Respectively, $t = 0$ represents the market opening.

Third, the average cumulative return (ACR) at time t :

$$ACR_t = \frac{1}{N} \sum_{i=1}^N CR_{i,t}, \tag{23}$$

was computed for all event days. This figure is available for any minute t after the start of the trading day. N is defined as the total number of days fulfilling the event day properties.

Fourth, t -tests were conducted to determine whether a given price movement after a specified event was significant. Specifically, we calculated the corresponding test statistic to examine if the ACR_t at time t was significantly distinct from zero. The test statistic had the following form:

$$t_{ACR_t} = \sqrt{N} \frac{(\overline{ACR}_t - 0)}{S_{ACR_t}} \sim t(N - 1), \tag{24}$$

where $0 < t \leq T$ and \overline{ACR}_t denotes the mean of the sample. Furthermore, S_{ACR_t} represents its standard deviation, and N defines the total numbers of days in the filtered dataset. Under the null hypothesis of no distinction from zero, the test statistic follows a t -distribution with $N - 1$ degrees of freedom.

Table 1 shows the characteristics of the overnight price gaps detected by our jump test procedure. In total, we observed 2128 overnight gaps during the sample period: 1154 of those gaps were positive, while 974 were negative. On average, the S&P 500 index faced positive (negative) overnight gaps of 0.60 percent (−0.67 percent). The largest overnight gaps occurred during the global financial crisis with 6.02 percent and −7.64 percent. The fact that both the range and the standard deviation of negative gaps were higher than those of positive overnight movements confirms the existing literature: market participants tend to react stronger to bad news rather than to good headlines (Suleman 2012). Concluding, Table 1 shows that there was a sufficient number of overnight price gaps leading to temporary market inefficiencies. As a result, this jump behavior generated high-frequency stock price dynamics that created major trading opportunities. In stark contrast to the approach of Fung et al. (2000) and Grant et al. (2005), the gaps identified by our jump-test scheme were both flexible and data-driven.

Table 1. Characteristics of positive and negative overnight gaps, which are identified by the Barndorff–Nielsen and Shephard (BNS) jump test, from January 1998–December 2015.

	Positive Gap	Negative Gap
Number of gaps	1154	974
Mean	0.0060	−0.0067
Minimum	0.0003	−0.0764
Quartile 1	0.0029	−0.0085
Median	0.0045	−0.0049
Quartile 3	0.0072	−0.0029
Maximum	0.0602	−0.0005
Standard deviation	0.0053	0.0063
Skewness	3.2771	−3.8289
Kurtosis	20.8453	29.3100

Figure 1 illustrates the detected jumps in a more detailed way. We observe a higher variation of negative overnight gaps, which is not surprising since financial data possess an asymmetric distribution (Cont 2001). Interestingly, the interval with the highest number of observations for both positive and negative overnight gaps was about ± 0.15 .

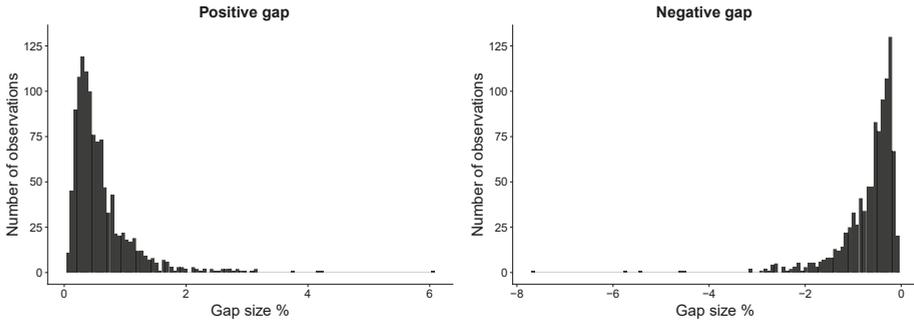


Figure 1. Histogram of positive and negative overnight gaps, which were identified by the BNS jump test, from January 1998–December 2015.

Figure 2 presents the number of detected overnight gaps over time. With rising volatility in financial markets, the number of overnight gaps also increased; fluctuations in the market imply jumps. Thus, it is not surprising that we observed almost no jumps in the first years of our sample period. In stark contrast, the number of overnight price gaps increased in times of high market turmoil. In general, more positive than negative gaps affect the S&P 500 index. As expected, this pattern changes during crises such as the dot-com crash in the early 2000s and the financial crisis in 2008. This also demonstrates the flexibility of the approach used to identify overnight gaps.

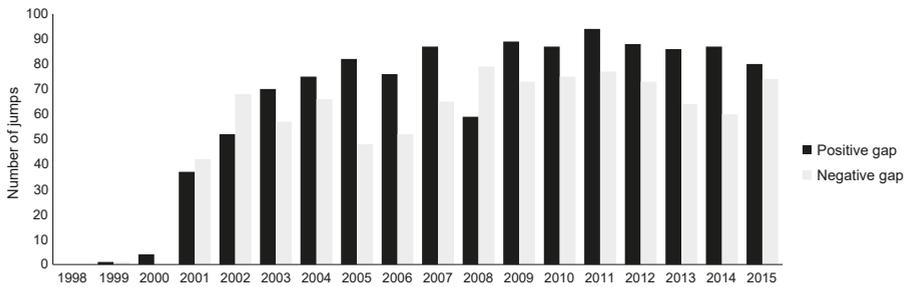


Figure 2. Development of positive and negative overnight gaps, which were identified by the BNS jump test, from 1998–2015.

Figure 3 depicts the average cumulative returns after overnight gaps identified by the BNS jump test. The detailed development of the ACR for positive and negative price gaps is reported in Table A1. The typical price pattern after overnight gaps is still persistent in modern financial markets, despite that markets should become more efficient in the course of digitalization and improved information flow (see Fung et al. (2000) and Grant et al. (2005)). In the case of a positive overnight gap, the average cumulative returns rose for a brief period before reverting to the minimum at -0.0316 percent. After reaching the lowest ACR 105 min after market opening, it began to rise until it crossed the zero percent line. From this point, the returns almost fell close to the minimum before increasing again. The upswing accelerated towards market closing, reaching 0.0236 percent at the end of the trading day. Following a negative overnight gap, the ACR move inverted. Starting with a brief continuation of the initial overnight movement, which marked the minimum of -0.0093 percent two

minutes after the stock exchange opens, the *ACR* began to reverse to its maximum of 0.0463 percent after approximately one and a half hours. The *ACR* remained relatively stable between 0.0200 and 0.0400 percent subsequent to hitting the upper limit. During the last ten minutes, the *ACR* rapidly decreased until the end of the trading day. Noticeable is that the magnitude of the variation of the *ACR* was stronger after negative price gaps. This is in line with stronger expected reactions of market participants to bad information that was also observable in the represented gap characteristics (Table 1). The *p*-values for both *ACR* realizations indicated that the returns were statistically different from zero on a 10 percent significance level for most of the time before the 115-min mark. After that threshold has passed, *p*-values well exceeded 10 percent; this fact is not surprising since many professional day traders stop trading after two trading hours because volatility and volume tend to decrease (see [Balance \(2019\)](#)). Furthermore, we recognized that the *ACR* for positive overnight gaps were not significant for a target time of 5, 35, 65, and 95 min based on a 10% significance level; it seems that the pattern is systematically repeated at 30-min intervals. This statement is confirmed by [Business Insider \(2015\)](#), which shows that the trading volume increases in the first minutes of every trading hour. Furthermore, [Bedowska-Sojka \(2013\)](#) demonstrated that this volatility is influenced by macroeconomic releases, which are typically published at 9:30, 10:00, 10:30, and 11:00. As a result, the test-statistic decreased, leading to non-significant *p*-values.

Concluding, our event study confirms the overreaction hypothesis and supports the results of [Fung et al. \(2000\)](#) and [Grant et al. \(2005\)](#). The findings of the event study further suggest that we are in a position to develop a statistical arbitrage strategy that exploits the mean-reversion characteristic of stocks after statistically-significant overnight price gaps (see [Poterba and Summers \(1988\)](#), [Leung and Li \(2015\)](#), [Lubnau and Todorova \(2015\)](#)). Specifically, it seems profitable to open trades after overnight gaps and close them after 2 h, i.e., we should set a target time of 120 min.

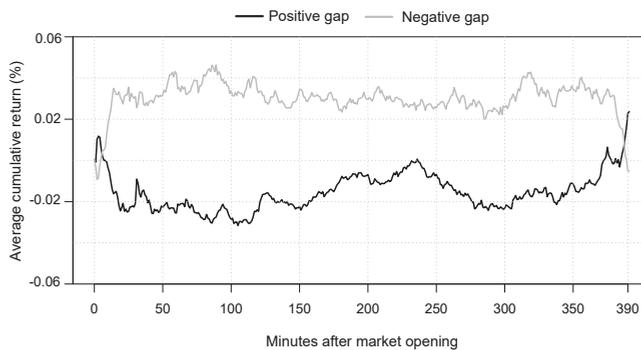


Figure 3. Average cumulative returns (%) after positive and negative overnight gaps, which were identified by the BNS jump test, from January 1998–December 2015.

4. Back-Testing Framework

The empirical back-testing study was performed from January 1998–December 2015 at intraday prices for the S&P 500 index components (see Section 4.1). According to [Gatev et al. \(2006\)](#) and [Nakajima \(2019\)](#), we divided the dataset into overlapping study periods, which were shifted by one day each. Each study period consisted of two consecutive phases. In the formation period (Section 4.2), the most appropriate stocks were selected using predefined models and criteria. In the subsequent out-of-sample trading period (Section 4.3), the top stocks were traded using rule-based entry and exit signals; this procedure avoids any look-ahead bias. Summarizing, we developed a full-fledged statistical arbitrage framework based on a jump–diffusion model (JDS), which is able to capture intraday and overnight high-frequency price dynamics.

4.1. Data and Software

The empirical back-testing was based on intraday data from the S&P 500 from January 1998–December 2015. This highly liquid stock market includes the stocks of the 500 leading blue chip companies that offer high-quality commodities and generally-accepted services. Since the S&P 500 index captures 80 percent of the total U.S. market capitalization ([S&P Dow Jones Indices 2015](#)), this dataset represents a fundamental test for any potential capital market anomaly. To be in line with [Stübinger and Endres \(2018\)](#), we applied a two-step process with the objective of removing any survivor bias from the database. First, we used the information list from [QuantQuote \(2016\)](#) to build a binary constituent matrix for S&P 500 shares from January 1998–December 2015. The 4527 rows characterize the trading days considered, and the 984 columns show the stocks that were ever in the S&P 500. Each element of this matrix displays a “1” if the corresponding company is part of the S&P 500 index on the corresponding day, otherwise a “0”. The sum of each row is about 500 because on each trading day, there are approximately 500 stocks in the index. Second, the complete archive of minute-by-minute prices from January 1998–December 2015 was downloaded from [QuantQuote \(2016\)](#). The corresponding stock exchange was open from Monday to Friday from 9:30–16:00 Eastern time. Consequently, the price time series of a share includes 391 data points per day. We followed [Stübinger and Endres \(2018\)](#) and adjusted the data by stock splits, dividends, and other corporate actions. By performing these two steps, our study design is in a position to map the constituents of the S&P 500 and the corresponding price time series completely.

The presented methodology and all relevant evaluations were implemented in the statistical programming language R ([R Core Team 2019](#)). For computation-intensive calculations, we used both the general-purpose programming language C++ and on-demand cloud computing platforms with virtual computer clusters that are available 24/7 via the Internet.

4.2. Formation Period

In the formation period, we considered all S&P 500 stock constituents. Therefore, we (i) conducted the BNS jump test based on past returns (ii) applied the jump detection scheme in the case of rejecting the null hypothesis, and (iii) selected the top stocks for the subsequent trading period. This subsection describes the outlined three-step logic.

In the first step, we executed the BNS jump test based on both the 390 intraday returns of the last trading day and the overnight return, i.e., the percentage change of the price from 16:00 of the last day to 9:30 of the current trading day. Specifically, we determined the z-statistic of [Huang and Tauchen \(2005\)](#) (see Equation (18)). If the null hypothesis was rejected, at least one jump emerged in the underlying security during the considered period. If the null hypothesis was not rejected, no jump emerged in the underlying security during the considered period. Consequently, we did not consider this stock in our back-testing framework.

In the second step, we applied the jump identification method of [Andersen et al. \(2010\)](#) to ensure that we only selected stocks possessing overnight gaps (see Section 2.2). Therefore, we considered only stocks that incorporate a significant overnight gap.

In the third step, we followed [Miao \(2014\)](#) and [Stübinger and Endres \(2018\)](#) and selected the most suitable shares for the out-of-sample trading period. Our algorithm attempted to find stocks possessing the most meaningful jump last night. For this purpose, we selected the top stocks, that possesses overnight gaps in the sense of [Andersen et al. \(2010\)](#), with the highest z-statistic of [Huang and Tauchen \(2005\)](#). The top 10 stocks were transferred to the trading period (see Section 4.3)¹.

¹ If less than 10 shares satisfied the condition of [Andersen et al. \(2010\)](#), we traded accordingly less. However, this case is extremely rare.

4.3. Trading Period

The top 10 stocks with the highest z-statistic were considered in the one-day trading period. For every top stock, we applied the following trading rules:

- We observe a negative price gap during the night, i.e., the stock is undervalued. Consequently, we go long in the stock.
- We observe a positive price gap during the night, i.e., the stock is overvalued. Consequently, we go short in the stock.

Motivated by Section 3, the trade was reversed 120 min later. Our strategy was based on a two-stage logic. First, we identified significant overnight price changes that had a substantial impact on future stock prices. Second, the top stocks possessed mean-reverting price dynamics, so that we could take advantage of these temporary market inefficiencies. If our assumption was correct, we were in a position to capture transient mispricings and generate profits. Concluding, we created a statistical arbitrage strategy based on a mean-reverting jump-diffusion model, the individual jump threshold depends on the underlying volatility.

As we aim for a classic long-short investment strategy in the sense of Gatev et al. (2006), we followed the principles of Avellaneda and Lee (2010) and Stübinger et al. (2018) and secured the market exposure with appropriate capital investments in the S&P 500 index. Every activity carried out on the market involves transaction costs. Therefore, it would be naive to ignore these fees as our high-frequency framework is based on permanent trading. According to Prager et al. (2012) and Stübinger and Bredthauer (2017), estimating exact values is not possible, but the bid-ask spread had abated to lower than one percent for stocks of the S&P 500 index, i.e., two basis points for an average stock price of 50 USD. In the same vein, Voya Investment Management (2016) accounted for a bid-ask spread of 3.5 basis points for the S&P 500, which was caused by increased use of algorithmic trading, decimalization, and changes in the stock market landscape. To be in line with Stübinger and Endres (2018), we assumed transaction costs of five basis points per share per half-turn. Consequently, transaction costs per complete round-trip corresponded to 20 basis points. This assumption appears realistic in light of our high-turnover strategy in a highly-liquid equity market.

In order to evaluate the value-add of our strategy, we benchmarked it against strategies from the same research field, but less flexible. More specifically, we considered the S&P 500 buy-and-hold strategy (BHS), fixed threshold strategy (FTS), general volatility strategy (GVS), and reverting volatility strategy (RVS) (see Table 2). The characteristic “individual” implies that the trading behavior depends on the underlying variable. If the model captured the behavior of fluctuations of stock price dynamics, we assigned the “volatility” property. The feature “mean-reverting” was fulfilled for statistical approaches that were able to model convergence to equilibrium after divergence. Finally, the explicit inclusion of a jump term led to the characteristic “jump-diffusion”. Data and the general frame were set identically to the JDS in order to ensure a fair comparison. Especially, we transferred the top 10 stocks to the trading period for each day across all strategies. Details of the four benchmark strategies are presented in the following paragraphs.

Table 2. Overview of the characteristics of the S&P 500 buy-and-hold strategy (BHS), fixed threshold strategy (FTS), generalized volatility strategy (GVS), reverting volatility strategy (RVS), and jump-diffusion strategy (JDS).

Characteristic	BHS	FTS	GVS	RVS	JDS
Individual	No	Yes	Yes	Yes	Yes
Volatility	No	No	Yes	Yes	Yes
Mean-reverting	No	No	No	Yes	Yes
Jump-diffusion	No	No	No	No	Yes

S&P 500 Buy-and-Hold Strategy (BHS)

First, we compared JDS to a naive S&P 500 buy-and-hold strategy (BHS). To be more specific, the index was bought in January 1998 and held during the complete time period. This passive investment neglected all the characteristics required for a successful strategy, namely, “individual”, “volatility”, “mean-reverting”, and “jump-diffusion”.

Fixed Threshold Strategy (FTS)

According to [Fung et al. \(2000\)](#), [Grant et al. \(2005\)](#), and [Caporale and Plastun \(2017\)](#), the fixed threshold strategy (FTS) detects abnormal overnight changes using a fixed threshold of ± 0.20 percent. This benchmark strategy obtains an individual trading limit for each stock. In our framework, the top 10 stocks with the highest absolute changes were opened at 9:30 of the trading day. We went long in the undervalued stocks and went short in the overvalued stocks. Identical to JDS, the positions were reversed 120 min after market opening. This approach was not in a position to distinguish stocks on the basis of their fluctuation behavior.

General Volatility Strategy (GVS)

The general volatility strategy (GVS) is based on the assumption that equities with high volatility exhibit temporary market inefficiencies (see [Banerjee et al. \(2007\)](#), [Barviera \(2017\)](#)). Following [Stübinger and Endres \(2018\)](#), we calculated the standard deviation of the overnight returns of the last 40 days and transferred the top 10 stocks with the highest volatility to the trading period. Again, undervalued (overvalued) stocks were bought (sold), and trades were reversed after 120 min.

Reverting Volatility Strategy (RVS)

Last but not least, the reverting volatility strategy (RVS) adds the mean-reversion component to GVS, i.e., we measured the degree of reversion to the equilibrium level after divergences. According to [Do and Faff \(2010\)](#), we determined the mean-reversion speed by the number of zero-crossings, which is defined as the number of times prices cross the zero line. Stocks were ranked separately by standard deviation and zero crossings; the stock with the highest value was assigned the highest rank for each measurement. Next, we formed a combined rank by the sum of the two separate ranks. The top 10 stocks were received by selecting stocks with the highest overall rank. The main disadvantage of this approach was the lack of a jump term, which reflects uncertainty in addition to the volatility component ([Cartea et al. 2015](#)).

5. Results

Following the high-frequency research studies of [Mitchell \(2010\)](#) and [Knoll et al. \(2018\)](#), we conducted a fully-fledged performance evaluation for the top 10 stocks of JDS from January 1998–December 2015 compared to the benchmarks BHS, FTS, GVS, and RVS. In particular, we evaluated the return characteristics and risk metrics (Section 5.1), examined the performance over time (Section 5.2), and analyzed the robustness of the strategies (Section 5.3). According to [Gatev et al. \(2006\)](#) and [Avellaneda and Lee \(2010\)](#), this paper calculated the total return based on committed capital, i.e., we divided the sum of daily net profits at the current day by the deployed capital.

5.1. Risk-Return Characteristics

Table 3 shows the daily return characteristics and risk metrics before and after transaction costs for the top 10 stocks per strategy from January 1998–December 2015. We observed statistically-significant returns for FTS, GVS, RVS, and JDS with Newey–West (NW) *t*-statistics above 15 prior to transaction costs. From an economical point of view, daily returns ranged between 0.17 percent for FTS and 0.36 percent for JDS. If we considered transaction costs, only the mean-reverting strategies RVS and JDS produced positively significant daily returns of 0.13 percent (RVS) and 0.17 percent (JDS). As

expected, BHS generated statistically non-significant returns of 0.02 percent per day (see [Endres and Stübinger \(2019b\)](#)). The range, i.e., the difference of the maximum and minimum, was vastly different for JDS (approximately 0.30 percentage points), compared to BHS, FTS, GVS, and RVS (approximately 0.15 percentage points); this dissimilarity is potentially driven by the jump-diffusion term. The same argument explains the increased standard deviation of JDS. All individual strategy variants depicted favorable characteristics for any potential investor due to the fact that the underlying returns showed right skewness and followed a leptokurtic distribution ([Cont 2001](#)). We found that the maximum drawdown was quite different for FTS (87.84 percent) and GVS (89.47 percent), in contrast to RVS (55.91 percent), BHS (64.33 percent), and JDS (68.17 percent); the difference between non-reverting and reverting top stocks is clearly pointed out. The hit rate of JDS, i.e., the percentage of days with non-negative returns, outperformed with 58.41 percent after transactions costs, compared to the benchmarks, ranging between 41.79 percent for FTS and 55.92 percent for RVS.

Table 3. Daily return characteristics and risk metrics for BHS, FTS, GVS, RVS, and JDS from January 1998–December 2015. NW denotes Newey–West standard errors with 1-lag correction and CVaR the conditional value at risk.

	Before Transaction Costs					After Transaction Costs			
	BHS	FTS	GVS	RVS	JDS	FTS	GVS	RVS	JDS
Mean return	0.0002	0.0017	0.0019	0.0033	0.0036	−0.0003	−0.0001	0.0013	0.0017
Standard error (NW)	0.0002	0.0001	0.0001	0.0001	0.0002	0.0001	0.0001	0.0001	0.0002
<i>t</i> -Statistic (NW)	0.8617	17.9433	15.8454	23.0251	16.4912	−2.5816	−1.1504	9.2534	7.8870
Minimum	−0.0947	−0.0410	−0.0521	−0.0544	−0.1169	−0.0430	−0.0541	−0.0564	−0.1187
Quartile 1	−0.0056	−0.0012	−0.0016	−0.0013	−0.0021	−0.0032	−0.0036	−0.0033	−0.0041
Median	0.0005	0.0012	0.0013	0.0030	0.0028	−0.0008	−0.0007	0.0010	0.0008
Quartile 3	0.0061	0.0040	0.0046	0.0076	0.0085	0.0020	0.0026	0.0056	0.0065
Maximum	0.1096	0.0604	0.0776	0.0889	0.1947	0.0584	0.0756	0.0869	0.1923
Standard deviation	0.0126	0.0062	0.0077	0.0090	0.0129	0.0062	0.0077	0.0090	0.0128
Skewness	−0.1987	1.2552	1.2987	0.9082	2.7078	1.2552	1.2987	0.9082	2.6990
Kurtosis	7.5278	9.5525	11.7337	8.3119	29.7136	9.5525	11.7337	8.3119	29.8425
Historical VaR 1%	−0.0350	−0.0136	−0.0178	−0.0187	−0.0255	−0.0156	−0.0198	−0.0207	−0.0275
Historical CVaR 1%	−0.0506	−0.0186	−0.0249	−0.0263	−0.0346	−0.0206	−0.0269	−0.0283	−0.0365
Historical VaR 5%	−0.0197	−0.0068	−0.0078	−0.0093	−0.0129	−0.0088	−0.0098	−0.0113	−0.0149
Historical CVaR 5%	−0.0302	−0.0110	−0.0141	−0.0155	−0.0209	−0.0130	−0.0161	−0.0175	−0.0228
Maximum drawdown	0.6433	0.0667	0.0860	0.1012	0.2707	0.8784	0.8947	0.5991	0.6817
Share with return ≥ 0	0.5313	0.6327	0.6200	0.6782	0.6715	0.4179	0.4288	0.5592	0.5841

In [Table 4](#), we depict annualized risk-return measures before transaction costs (left side) and after transaction costs (right side). After transaction costs, JDS produced returns of 51.47 percent p.a., compared to 38.85 percent for RVS, −4.07 percent for GVS, and −6.59 percent for FTS. Thus, the first two strategies achieved meaningfully better results than the naive buy-and-hold strategy (BHS) with an average return of 1.81 percent p.a. Across all strategies, the mean excess return was similar to the mean return because the risk-free rate was very close to zero, especially in the last years. Our jump-based strategy JDS generated approximately the standard deviation of the market, resulting in a Sharpe ratio of 2.38 after transaction costs. This value confirmed the results of the high-frequency studies of [Knoll et al. \(2018\)](#) and [Stübinger \(2018\)](#). The lower partial moment risk of JDS led to a Sortino ratio of 4.76, compared to the benchmarks ranging between −1.03 (FTS) and 4.67 (RVS). We summarized that JDS outperformed the classic approaches in a large number of comparisons; complexity pays off. Our task was still to evaluate the performance over time, as well as the robustness of the strategies.

Table 4. Annualized risk-return measures for BHS, FTS, GVS, RVS, and JDS from January 1998–December 2015.

	Before Transaction Costs					After Transaction Costs			
	BHS	FTS	GVS	RVS	JDS	FTS	GVS	RVS	JDS
Mean return	0.0181	0.5456	0.5874	1.2959	1.4472	−0.0659	−0.0407	0.3885	0.5147
Mean excess return	−0.0022	0.5149	0.5558	1.2503	1.3985	−0.0846	−0.0598	0.3609	0.4845
Standard deviation	0.2005	0.0984	0.1219	0.1432	0.2045	0.0984	0.1219	0.1432	0.2037
Downside deviation	0.1441	0.0490	0.0633	0.0696	0.0950	0.0639	0.0777	0.0832	0.1082
Sharpe ratio	−0.0110	5.2312	4.5598	8.7339	6.8392	−0.8592	−0.4904	2.5211	2.3781
Sortino ratio	0.1256	11.1380	9.2757	18.6058	15.2388	−1.0315	−0.5229	4.6719	4.7587

5.2. Sub-Period Analysis

Motivated by the time-varying returns of Liu et al. (2017) and Stübinger and Knoll (2018), we analyzed the stability and potential of the strategies over time. Figure 4, therefore, presents the development of an investment of USD 1 after transaction costs for FTS, GVS, RVS, JDS (first column), and the S&P 500 buy-and-hold strategy BHS (second column) over three partial periods. Table A2 provides a detailed overview of the corresponding annualized risk-return ratios for sub-periods of three years.

The first sub-period ran from 1998–2006 and described the bursting of the Internet bubble and the start of the Iraq war, as well as the subsequent bull market. We observed meaningful differences in performance between the mean-reverting and non-mean-reverting strategies: the average annual returns after transaction costs of up to 73.76 percent for RVS and up to 64.08 percent for JDS were well above those of BHS (7.87 percent), FTS (27.31 percent), and GVS (42.26 percent). As a typical feature in the financial context, the baseline methods were nevertheless successful in this period due to market inefficiencies and a lack of transparency.

The second sub-period ranged from 2007–2009 and was characterized by the global financial crisis and its consequences. In the course of the sub-prime crisis, the overall market showed strong fluctuations and substantial declines. In contrast, the other strategies generated positive returns, ranging from 27.35 percent for FTS to 315.02 percent for JDS. This strong performance was not astonishing as Avellaneda and Lee (2010) and Rad et al. (2016) demonstrated that statistical arbitrage trading strategies achieved abnormal returns during bear markets.

The third sub-period extends from 2010–2015 and covered a period of comebacks and restarts. The benchmarks FTS and GVS showed declining trends compared to the overall market, caused by the increasing public availability of these methods. RVS achieved an almost constant cumulative return of one, i.e., this strategy generated exactly the costs that were incurred. For JDS, we observed that 1 USD invested in January 2010 grew to 5 USD after transaction costs; performance did not decline across time and seemed to be robust against drawdowns.

5.3. Robustness Check

As mentioned above, we motivated the target time of 120 min based both on the available literature and the results of our event study; see Section 3. Since data snooping is a major problem in many financial applications, this subsection examines the sensitivity of our strategies to deviations from their parameter value. In Table 5, we vary the target time in two directions and report the annualized returns before and after transaction costs for BHS, FTS, GVS, RVS, and JDS.

First of all, we see that our results were robust in the face of parameter variations and always led to statements similar to those in Section 5.1. As expected, the results of a target time of 120 were identical to those of Table 3. Furthermore, the annualized returns for each strategy converged as the relative change decreased with increasing target time. The naive S&P 500 buy-and-hold strategy (BHS) always led to an annualized return of 1.81 percent, which is not surprising, since this approach is completely independent of the target time (Section 4). Furthermore, the performance of FTS increased slightly with ascending target time, e.g., the annualized return after transaction costs was −9.37 percent

if we closed the trade at 9:50 and -8.36 percent if we closed it at 13:10. The same statement applies to GVS (-9.70 percent vs. -4.28 percent). Due to their mean-reverting component, RVS and JDS showed a slightly declining performance. For each target time, JDS remained the best variant with annualized returns between 49.65 percent and 62.61 percent, after transaction costs. Obviously, we were not on an optimum, but we found robust trading results, regardless of fluctuations in our parameter setting.

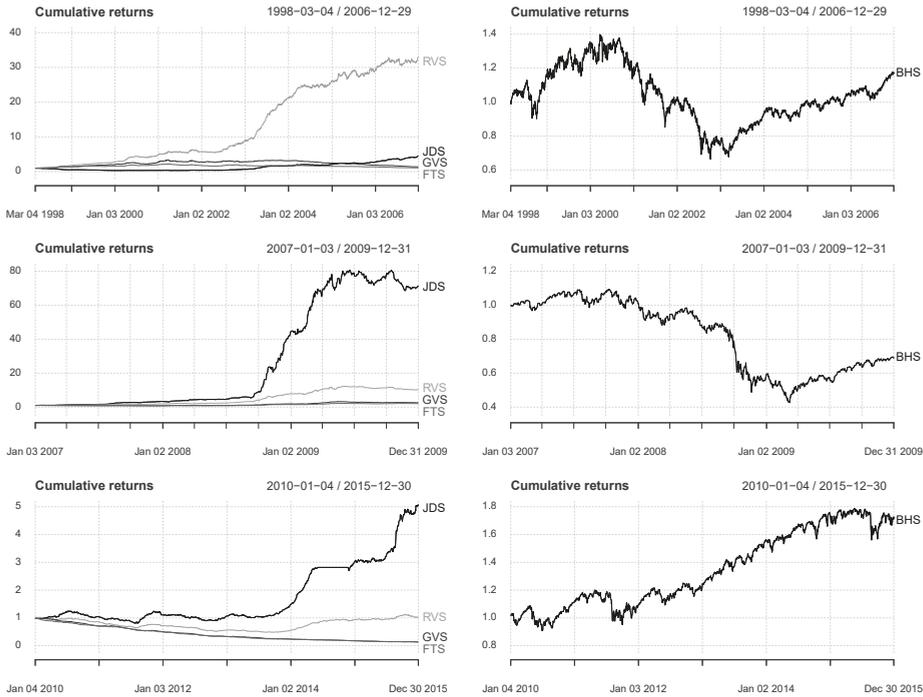


Figure 4. Development of an investment of 1 USD after transaction costs for FTS, GVS, RVS, and JDS (first column) compared to the S&P 500 buy-and-hold-strategy (BHS) (second column). The time period from January 1998–December 2015 is divided into three sub-periods (March 1998/December 2006, January 2007/December 2009, January 2010/December 2015).

Table 5. Yearly returns for BHS, FTS, GVS, RVS, and JDS for a varying target time from January 1998–December 2015.

Target Time	Before Transaction Costs					After Transaction Costs			
	BHS	FTS	GVS	RVS	JDS	FTS	GVS	RVS	JDS
20 min	0.0181	0.4997	0.4944	1.4201	1.6941	-0.0937	-0.0970	0.4638	0.6261
40 min	0.0181	0.5030	0.5382	1.4120	1.6685	-0.0918	-0.0704	0.4589	0.6104
60 min	0.0181	0.5214	0.5525	1.3706	1.6624	-0.0806	-0.0618	0.4338	0.6068
80 min	0.0181	0.5088	0.5483	1.3132	1.5883	-0.0882	-0.0643	0.3990	0.5628
100 min	0.0181	0.5065	0.5583	1.3107	1.5893	-0.0896	-0.0583	0.3975	0.5634
120 min	0.0181	0.5456	0.5874	1.2959	1.4472	-0.0659	-0.0407	0.3885	0.5147
140 min	0.0181	0.5346	0.5748	1.2748	1.5233	-0.0726	-0.0483	0.3757	0.5241
160 min	0.0181	0.5384	0.5897	1.2653	1.5226	-0.0703	-0.0392	0.3700	0.5229
180 min	0.0181	0.5699	0.5848	1.2510	1.4946	-0.0512	-0.0422	0.3613	0.5061
200 min	0.0181	0.5268	0.5764	1.2255	1.4783	-0.0773	-0.0473	0.3459	0.4965
220 min	0.0181	0.5165	0.5838	1.2358	1.4865	-0.0836	-0.0428	0.3521	0.5014

Motivated by the findings in Section 3, Table 6 examines the annualized returns for a target time of 5, 35, 65, and 95 min. Most interestingly, annual returns were substantially lower for a target time of 5 min for FTS, GVS, RVS, and JDS because high market turmoil during the opening minutes reduced the results. For a target time of 35, 65, and 95 min, increasing market efficiency during the first minutes of each trading hour did not affect yearly returns before and after transaction costs; our strategies seem to be robust against this effect.

Table 6. Yearly returns for BHS, FTS, GVS, RVS, and JDS for a target time of 5, 35, 65, and 95 min from January 1998–December 2015.

Target Time	Before Transaction Costs					After Transaction Costs			
	BHS	FTS	GVS	RVS	JDS	FTS	GVS	RVS	JDS
5 min	0.0181	0.3217	0.2529	1.0387	1.1842	−0.2015	−0.2432	0.2327	0.3174
35 min	0.0181	0.5233	0.5349	1.4023	1.6719	−0.0795	−0.0725	0.4531	0.6134
65 min	0.0181	0.5232	0.5515	1.3793	1.6431	−0.0795	−0.0624	0.4391	0.5956
95 min	0.0181	0.5118	0.5601	1.3102	1.5824	−0.0864	−0.0572	0.3972	0.5589

Next, we take a closer look at our S&P 500 buy-and-hold strategy (BHS). The S&P 500 index was purchased in January 1998 and was held for the entire sample period. Of course, BHS is only a baseline approach for betting on the market. Therefore, we followed [Endres and Stübinger \(2019b\)](#) and developed a more realistic benchmark: The S&P 500 strategy buys the index at 9:30 and reverses it after 120 min. We observed an annualized return of 1.03% compared to 1.81% for BHS (see also Table 4). This insufficient performance is not surprising, as it is a baseline approach without modeling.

Finally, this manuscript supposed a high-turnover strategy of an institutional trader on high-frequency prices. Motivated by the literature, our back-testing framework assumed transaction costs of five basis points per share per half-turn, resulting in 20 basis points per round-trip per pair. However, other traders may be less aggressive in implementing this strategy. Therefore, we analyzed the breakeven point of the statistical arbitrage strategy since investors are exposed to different market conditions. We found that the breakeven point of JDS was between 35 basis points and 40 basis points. Concluding, this strategy generated promising results, even for investors that are exposed to different market conditions and thus higher transaction costs.

6. Conclusions

In this paper, we presented an integrated statistical arbitrage strategy based on overnight price gaps and implemented it on high-frequency data of the S&P 500 stocks from January 1998–December 2015. In this context, we made four contributions to the literature. The first contribution relates to the developed trading framework based on a jump-diffusion model: we are in a position to capture jumps, mean-reversion, volatility clusters, and drifts. Our approach identifies overnight price gaps based on the jump test of [Barndorff-Nielsen and Shephard \(2004\)](#) and exploits temporary market anomalies by corresponding investments. In a preliminary study, we confirmed the assumption of mean-reverting overnight gaps with the aid of the S&P 500 index. The second contribution focuses on the value-add of our strategy. Therefore, we benchmarked it against well-known quantitative strategies from the same research area, namely the naive S&P 500 buy-and-hold strategy, fixed threshold strategy, general volatility strategy, and reverting volatility strategy. The third contribution is based on our large-scale empirical study on a sophisticated back-testing framework. Our strategy produced statistically- and economically-significant returns of 51.47 percent p.a. after transaction costs; the benchmarks were outperformed. The fourth contribution focuses on the profitable and robust performance results also in the last part of our sample period. Our findings posited a severe challenge to the semi-strong form of market efficiency even in recent times.

We identified three possible directions for further research: First, the event study and the back-testing framework should be conducted in other equity universes. Second, the exit signal

of the strategy should be determined for each stock individually. Third, a multivariate model could be developed that takes into account the common interactions between stocks.

Author Contributions: J.S. and L.S. conceived of the research method. The experiments were designed by J.S. and performed by L.S. The analyses were conducted and reviewed by J.S. and L.S. The paper was initially drafted by J.S. and revised by L.S. It was refined and finalized by L.S. and J.S.

Funding: We are grateful to the “Open Access Publikationsfonds”, which has covered 75 percent of the publication fees.

Acknowledgments: We are further grateful to Ingo Klein and two anonymous referees for many helpful discussions and suggestions on this topic.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Detailed development of the ACR with *p*-values of the two-sided *t*-test from January 1998–December 2015. ACR denotes the average cumulative returns.

Target Time	Positive Gap		Negative Gap	
	ACR in %	<i>p</i> -Value	ACR in %	<i>p</i> -Value
5 min	0.0056	0.1180	0.0013	0.8140
10 min	−0.0037	0.4690	0.0178	0.0180
15 min	−0.0155	0.0160	0.0334	0.0000
20 min	−0.0229	0.0020	0.0293	0.0060
25 min	−0.0248	0.0030	0.0355	0.0030
30 min	−0.0207	0.0260	0.0310	0.0160
35 min	−0.0150	0.1580	0.0327	0.0320
40 min	−0.0193	0.0850	0.0284	0.0670
45 min	−0.0245	0.0400	0.0314	0.0480
50 min	−0.0227	0.0670	0.0318	0.0560
55 min	−0.0236	0.0610	0.0415	0.0160
60 min	−0.0255	0.0490	0.0420	0.0150
65 min	−0.0209	0.1230	0.0346	0.0490
70 min	−0.0231	0.1000	0.0333	0.0690
75 min	−0.0257	0.0810	0.0387	0.0470
80 min	−0.0287	0.0600	0.0417	0.0310
85 min	−0.0301	0.0490	0.0432	0.0280
90 min	−0.0256	0.1030	0.0415	0.0360
95 min	−0.0230	0.1540	0.0373	0.0690
100 min	−0.0286	0.0810	0.0338	0.0990
105 min	−0.0316	0.0550	0.0334	0.1060
110 min	−0.0288	0.0820	0.0334	0.1070
115 min	−0.0294	0.0780	0.0382	0.0740
120 min	−0.0248	0.1360	0.0328	0.1260
130 min	−0.0189	0.2620	0.0307	0.1670
140 min	−0.0207	0.2290	0.0256	0.2660
150 min	−0.0224	0.2150	0.0343	0.1430
160 min	−0.0177	0.3300	0.0313	0.1920
170 min	−0.0156	0.3940	0.0295	0.2230
180 min	−0.0091	0.6250	0.0248	0.3160
190 min	−0.0066	0.7270	0.0276	0.2620
200 min	−0.0069	0.7180	0.0296	0.2320
210 min	−0.0111	0.5700	0.0340	0.1770
220 min	−0.0070	0.7210	0.0304	0.2370
230 min	−0.0027	0.8880	0.0258	0.3260
240 min	−0.0038	0.8450	0.0288	0.2650
250 min	−0.0068	0.7300	0.0254	0.3340
260 min	−0.0121	0.5460	0.0302	0.2640
270 min	−0.0175	0.3850	0.0314	0.2470
280 min	−0.0218	0.2890	0.0275	0.3260
290 min	−0.0212	0.3130	0.0241	0.3960
310 min	−0.0189	0.3880	0.0347	0.2390
330 min	−0.0131	0.5740	0.0309	0.3140
350 min	−0.0115	0.6360	0.0344	0.2820
370 min	−0.0073	0.7720	0.0276	0.4320
390 min	0.0229	0.4070	−0.0052	0.8900
391 min	0.0236	0.3900	−0.0052	0.8870

Table A2. Annualized risk-return measures for BHS, FTS, GVS, RVS, and JDS for sub-periods of 3 years from January 1998–December 2015.

		Before Transaction Costs					After Transaction Costs			
		BHS	FTS	GVS	RVS	JDS	FTS	GVS	RVS	JDS
1998–2000	Mean return	0.0624	1.1054	1.3520	1.8718	0.1674	0.2731	0.4226	0.7376	-0.2956
	Mean excess return	0.0106	1.0030	1.2378	1.7324	0.1105	0.2111	0.3534	0.6531	-0.3300
	Standard deviation	0.2055	0.1037	0.1172	0.1193	0.1439	0.1037	0.1172	0.1193	0.1435
	Downside deviation	0.1442	0.0449	0.0546	0.0515	0.1035	0.0591	0.0687	0.0641	0.1191
	Sharpe ratio	0.0516	9.6750	10.5590	14.5236	0.7682	2.0364	3.0143	5.4756	-2.2991
	Sortino ratio	0.4324	24.6231	24.7455	36.3203	1.6165	4.6206	6.1472	11.5012	-2.4817
2001–2003	Mean return	-0.0781	0.5919	0.7374	1.7218	1.5412	-0.0379	0.0502	0.6467	0.6408
	Mean excess return	-0.0978	0.5580	0.7005	1.6640	1.4872	-0.0584	0.0278	0.6116	0.6058
	Standard deviation	0.2184	0.1214	0.1463	0.1636	0.2037	0.1214	0.1463	0.1636	0.2024
	Downside deviation	0.1538	0.0685	0.0827	0.0823	0.1057	0.0849	0.0981	0.0966	0.1187
	Sharpe ratio	-0.4478	4.5978	4.7880	10.1681	7.2995	-0.4816	0.1902	3.7375	2.9939
	Sortino ratio	-0.5080	8.6396	8.9161	20.9326	14.5766	-0.4467	0.5116	6.6958	5.3993
2004–2006	Mean return	0.0787	0.4021	0.2554	0.9144	1.3252	-0.1529	-0.2417	0.1574	0.4035
	Mean excess return	0.0475	0.3616	0.2191	0.8592	1.2582	-0.1774	-0.2636	0.1240	0.3630
	Standard deviation	0.1046	0.0601	0.0656	0.1019	0.1278	0.0601	0.0656	0.1019	0.1276
	Downside deviation	0.0720	0.0323	0.0396	0.0499	0.0613	0.0483	0.0562	0.0650	0.0758
	Sharpe ratio	0.4542	6.0149	3.3407	8.4349	9.8430	-2.9506	-4.0187	1.2171	2.8453
	Sortino ratio	1.0935	12.4580	6.4486	18.3127	21.6319	-3.1672	-4.2977	2.4206	5.3199
2007–2009	Mean return	-0.1177	1.1060	1.2654	2.5881	5.8734	0.2735	0.3701	1.1720	3.1502
	Mean excess return	-0.1358	1.0628	1.2189	2.5147	5.7331	0.2473	0.3419	1.1274	3.0653
	Standard deviation	0.2995	0.1500	0.1991	0.2193	0.3477	0.1500	0.1991	0.2193	0.3470
	Downside deviation	0.2209	0.0687	0.0977	0.1009	0.1323	0.0823	0.1111	0.1134	0.1437
	Sharpe ratio	-0.4534	7.0874	6.1215	11.4693	16.4905	1.6491	1.7170	5.1421	8.8346
	Sortino ratio	-0.5328	16.0953	12.9488	25.6392	44.4096	3.3231	3.3318	10.3376	21.9222
2010–2012	Mean return	0.0671	0.1413	0.1445	0.3591	0.6918	-0.3107	-0.3088	-0.1789	0.0215
	Mean excess return	0.0663	0.1404	0.1436	0.3581	0.6905	-0.3112	-0.3093	-0.1795	0.0207
	Standard deviation	0.1856	0.0538	0.0651	0.1017	0.1543	0.0538	0.0651	0.1017	0.1540
	Downside deviation	0.1341	0.0328	0.0415	0.0658	0.0905	0.0512	0.0590	0.0815	0.1061
	Sharpe ratio	0.3572	2.6115	2.2080	3.5202	4.4759	-5.7880	-4.7542	-1.7647	0.1348
	Sortino ratio	0.5004	4.3142	3.4861	5.4583	7.6420	-6.0655	-5.2322	-2.1942	0.2029
2013–2015	Mean return	0.1219	0.2262	0.2275	1.0296	1.5703	-0.2593	-0.2585	0.2272	0.6838
	Mean excess return	0.1219	0.2262	0.2275	1.0296	1.5703	-0.2593	-0.2585	0.2272	0.6838
	Standard deviation	0.1281	0.0487	0.0611	0.1022	0.1392	0.0487	0.0611	0.1022	0.1372
	Downside deviation	0.0904	0.0289	0.0367	0.0506	0.0535	0.0455	0.0535	0.0647	0.0662
	Sharpe ratio	0.9516	4.6472	3.7216	10.0702	11.2826	-5.3274	-4.2295	2.2221	4.9853
	Sortino ratio	1.3484	7.8150	6.2034	20.3644	29.3617	-5.7035	-4.8294	3.5130	10.3278

References

Andersen, Torben G., and Tim Bollerslev. 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39: 885. [CrossRef]

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys. 2001. The distribution of realized exchange rate volatility. *Journal of the American Statistical Association* 96: 42–55. [CrossRef]

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys. 2003. Modeling and forecasting realized volatility. *Econometrica* 71: 579–625. [CrossRef]

Andersen, Torben G., Tim Bollerslev, Per Frederiksen, and Morten Ørregaard Nielsen. 2010. Continuous-time models, realized volatilities, and testable distributional implications for daily stock returns. *Journal of Applied Econometrics* 25: 233–61. [CrossRef]

Avellaneda, Marco, and Jeong-Hyun Lee. 2010. Statistical arbitrage in the US equities market. *Quantitative Finance* 10: 761–82. [CrossRef]

Balance. 2019. Make Money Personal. Available online: <https://www.thebalance.com> (accessed on 27 February 2019).

Banerjee, Prithviraj S., James S. Doran, and David R. Peterson. 2007. Implied volatility and future portfolio returns. *Journal of Banking & Finance* 31: 3183–99.

- Bariviera, Aurelio F. 2017. The inefficiency of bitcoin revisited: A dynamic approach. *Economics Letters* 161: 1–4. [CrossRef]
- Barndorff-Nielsen, Ole E., and Neil Shephard. 2002. Estimating quadratic variation using realized variance. *Journal of Applied Econometrics* 17: 457–77. [CrossRef]
- Barndorff-Nielsen, Ole E., and Neil Shephard. 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2: 1–37. [CrossRef]
- Barndorff-Nielsen, Ole E., and Neil Shephard. 2006. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics* 4: 1–30. [CrossRef]
- Bedowska-Sojka, Barbara. 2013. Macroeconomic news effects on the stock markets in intraday data. *Central European Journal of Economic Modelling and Econometrics* 5: 249–69.
- Bertram, William K. 2010. Analytic solutions for optimal statistical arbitrage trading. *Physica A: Statistical Mechanics and Its Applications* 389: 2234–43. [CrossRef]
- Business Insider. 2015. Markets Insider by Intelligence. Available online: <https://www.businessinsider.com/> (accessed on 27 February 2019).
- Caporale, Guglielmo Maria, and Alex Plastun. 2017. Price gaps: Another market anomaly? *Investment Analysts Journal* 46: 279–93. [CrossRef]
- Cartea, Álvaro, Sebastian Jaimungal, and José Penalva. 2015. *Algorithmic and High-Frequency Trading*. Cambridge: Cambridge University Press.
- Chen, Huafeng, Shaojun Chen, Zhuo Chen, and Feng Li. 2017. Empirical investigation of an equity pairs trading strategy. *Management Science* 65: 370–89. [CrossRef]
- Cont, Rama. 2001. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance* 1: 223–36. [CrossRef]
- Do, Binh, and Robert Faff. 2010. Does simple pairs trading still work? *Financial Analysts Journal* 66: 83–95. [CrossRef]
- Do, Binh, and Robert Faff. 2012. Are pairs trading profits robust to trading costs? *Journal of Financial Research* 35: 261–87. [CrossRef]
- Endres, Sylvia, and Johannes Stübinger. 2019a. Optimal trading strategies for Lévy-driven Ornstein-Uhlenbeck processes. *Applied Economics*, forthcoming. [CrossRef]
- Endres, Sylvia, and Johannes Stübinger. 2019b. Regime-switching modeling of high-frequency stock returns with Lévy jumps. *Quantitative Finance*, forthcoming. [CrossRef]
- Evans, Kevin P. 2011. Intraday jumps and us macroeconomic news announcements. *Journal of Banking & Finance* 35: 2511–27.
- Frömmel, Michael, Xing Han, and Frederick van Gysegem. 2015. Further evidence on foreign exchange jumps and news announcements. *Emerging Markets Finance and Trade* 51: 774–87. [CrossRef]
- Fung, Alexander Kwok-Wah, Debby MY Mok, and Kin Lam. 2000. Intraday price reversals for index futures in the US and Hong Kong. *Journal of Banking & Finance* 24: 1179–201.
- Gatev, Evan, William N. Goetzmann, and K. Geert Rouwenhorst. 2006. Pairs trading: Performance of a relative-value arbitrage rule. *Review of Financial Studies* 19: 797–827. [CrossRef]
- Göncü, Ahmet, and Erdinc Akyildirim. 2016. A stochastic model for commodity pairs trading. *Quantitative Finance* 16: 1843–57. [CrossRef]
- Grant, James L, Avner Wolf, and Susana Yu. 2005. Intraday price reversals in the US stock index futures market: A 15-year study. *Journal of Banking & Finance* 29: 1311–27.
- Hogan, Steve, Robert Jarro, Melvyn Teo, and Mitch Warachka. 2004. Testing market efficiency using statistical arbitrage with applications to momentum and value strategies. *Journal of Financial Economics* 73: 525–65. [CrossRef]
- Huang, Xin, and George Tauchen. 2005. The relative contribution of jumps to total price variance. *Journal of Financial Econometrics* 3: 456–99. [CrossRef]
- Knoll, Julian, Johannes Stübinger, and Michael Grottkke. 2018. Exploiting social media with higher-order factorization machines: Statistical arbitrage on high-frequency data of the S&P 500. *Quantitative Finance*, forthcoming.
- Larsson, Stig, Carl Lindberg, and Marcus Warfheimer. 2013. Optimal closing of a pair trade with a model containing jumps. *Applications of Mathematics* 58: 249–68. [CrossRef]

- Leung, Tim, and Xin Li. 2015. Optimal mean reversion trading with transaction costs and stop-loss exit. *International Journal of Theoretical and Applied Finance* 18: 1550020. [CrossRef]
- Liu, Bo, Lo-Bin Chang, and Hélyette Geman. 2017. Intraday pairs trading strategies on high frequency data: The case of oil companies. *Quantitative Finance* 17: 87–100. [CrossRef]
- Lubnau, Thorben, and Neda Todorova. 2015. Trading on mean-reversion in energy futures markets. *Energy Economics* 51: 312–19. [CrossRef]
- Miao, George J. 2014. High frequency and dynamic pairs trading based on statistical arbitrage using a two-stage correlation and cointegration approach. *International Journal of Economics and Finance* 6: 96–110. [CrossRef]
- Mitchell, John B. 2010. Soybean futures crush spread arbitrage: Trading strategies and market efficiency. *Journal of Risk and Financial Management* 3: 63–96. [CrossRef]
- Nakajima, Tadahiro. 2019. Expectations for statistical arbitrage in energy futures markets. *Journal of Risk and Financial Management* 12: 14. [CrossRef]
- Pole, Andrew. 2011. *Statistical Arbitrage: Algorithmic Trading Insights and Techniques*. Hoboken: John Wiley & Sons.
- Poterba, James M., and Lawrence H. Summers. 1988. Mean reversion in stock prices: Evidence and implications. *Journal of Financial Economics* 22: 27–59. [CrossRef]
- Prager, Richard, Supurna Vedbrat, Chris Vogel, and Even Cameron Watt. 2012. *Got Liquidity?* New York: BlackRock Investment Institute.
- QuantQuote. 2016. QuantQuote Market Data and Software. Available online: <https://quantquote.com> (accessed on 27 February 2019).
- R Core Team. 2019. *Stats: A Language and Environment for Statistical Computing*. R package. Wien: R Core Team.
- Rad, Hossein, Rand Kwong Yew Low, and Robert Faff. 2016. The profitability of pairs trading strategies: Distance, cointegration and copula methods. *Quantitative Finance* 16: 1541–58. [CrossRef]
- Rombouts, Jeroen V. K., and Lars Stentoft. 2011. Multivariate option pricing with time varying volatility and correlations. *Journal of Banking & Finance* 35: 2267–81.
- S&P Dow Jones Indices. 2015. S&P Global—Equity S&P 500 Index. Available online: <https://us.spindices.com/indices/equity/sp-500> (accessed on 27 February 2019).
- Stübinger, Johannes. 2018. Statistical arbitrage with optimal causal paths on high-frequency data of the S&P 500. *Quantitative Finance*, forthcoming.
- Stübinger, Johannes, and Jens Bredthauer. 2017. Statistical arbitrage pairs trading with high-frequency data. *International Journal of Economics and Financial Issues* 7: 650–62.
- Stübinger, Johannes, and Sylvia Endres. 2018. Pairs trading with a mean-reverting jump-diffusion model on high-frequency data. *Quantitative Finance* 18: 1735–51. [CrossRef]
- Stübinger, Johannes, and Julian Knoll. 2018. Beat the bookmaker—Winning football bets with machine learning (Best Application Paper). Paper presented at 38th SGAI International Conference on Artificial Intelligence, Cambridge, UK, December 11–13, pp. 219–33.
- Stübinger, Johannes, Benedikt Mangold, and Christopher Krauss. 2018. Statistical arbitrage with vine copulas. *Quantitative Finance* 18: 1831–49. [CrossRef]
- Suleman, Muhammad Tahir. 2012. Stock market reaction to good and bad political news. *Asian Journal of Finance & Accounting* 4: 299–312.
- Vidyamurthy, Ganapathy. 2004. *Pairs Trading: Quantitative Methods and Analysis*. Hoboken: John Wiley & Sons.
- Voya Investment Management. 2016. The Impact of Equity Market Fragmentation and Dark Pools on Trading and Alpha Generation. Available online: <https://investments.voya.com> (accessed on 27 February 2019).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Article

Between \mathbb{P} and \mathbb{Q} : The $\mathbb{P}^{\mathbb{Q}}$ Measure for Pricing in Asset Liability Management

Marcel T. P. van Dijk ^{1,2,*}, Cornelis S. L. de Graaf ¹ and Cornelis W. Oosterlee ^{2,3}

¹ Ortec Finance, 3011 XB Rotterdam, The Netherlands; csldegraaf@gmail.com

² DIAM—Delft Institute of Applied Mathematics, Delft University of Technology, 2628 CD Delft, The Netherlands; C.W.Oosterlee@cwi.nl

³ CWI—The Center for Mathematics and Computer Science, 1098 XG Amsterdam, The Netherlands

* Correspondence: marcel.vandijk@ortec-finance.com

Received: 17 September 2018; Accepted: 21 October 2018; Published: 24 October 2018

Abstract: Insurance companies issue guarantees that need to be valued according to the market expectations. By calibrating option pricing models to the available implied volatility surfaces, one deals with the so-called risk-neutral measure \mathbb{Q} , which can be used to generate market consistent values for these guarantees. For asset liability management, insurers also need future values of these guarantees. Next to that, new regulations require insurance companies to value their positions on a one-year horizon. As the option prices at $t = 1$ are unknown, it is common practice to assume that the parameters of these option pricing models are constant, i.e., the calibrated parameters from time $t = 0$ are also used to value the guarantees at $t = 1$. However, it is well-known that the parameters are not constant and may depend on the state of the market which evolves under the real-world measure \mathbb{P} . In this paper, we propose improved regression models that, given a set of market variables such as the VIX index and risk-free interest rates, estimate the calibrated parameters. When the market variables are included in a real-world simulation, one is able to assess the calibrated parameters (and consequently the implied volatility surface) in line with the simulated state of the market. By performing a regression, we are able to *predict* out-of-sample implied volatility surfaces accurately. Moreover, the impact on the *Solvency Capital Requirement* has been evaluated for different points in time. The impact depends on the initial state of the market and may vary between -46% and $+52\%$.

Keywords: insurance; Solvency II; risk-neutral models

1. Introduction

Liabilities of insurance companies depend on the fair value of the outstanding claims that typically involve guarantees (that are also called embedded options). The market consistent value of these guarantees is defined under the risk-neutral measure \mathbb{Q} , i.e., they are then computed with pricing formulas that agree on the current implied volatility surfaces. To hedge against the risks involved in these claims, insurers often acquire (complex) option portfolios that also require market consistent risk-neutral valuation themselves. Furthermore, on 1 January 2016 the so-called Solvency II directive came into effect which introduced the Solvency Capital Requirement (SCR). The SCR is defined as the minimum amount of capital which should be held by an insurer, such that the insurer is able to pay its claims over a one-year horizon with a 99.5% probability. The regulator demands that the insurer's available capital should be greater than, or equal to, the SCR. Because the claims typically depend on *future* market consistent valuation, computing the SCR, and, more generally performing proper Asset Liability Management (ALM), is a challenging task.

To compute these future market consistent values of the embedded options, insurers require the probability distribution of the values of these embedded options. Typically, this is done by simulating a large number of random future states of the market and, after that, the different states are valued

under the market consistent risk-neutral measure. From the simulated embedded option values, the desired statistics can then be extracted. The future states of the market can be computed by means of risk-neutral models (\mathbb{Q} in \mathbb{Q}), or real-world models (\mathbb{Q} in \mathbb{P}). Risk-neutral simulations are, for example, used to calculate the Credit Value Adjustment (CVA) (see, e.g., Pykhtin (2012)), which is a traded quantity and should therefore be computed using no-arbitrage arguments. For quantities that are not traded (or hedged), the \mathbb{Q} in \mathbb{Q} approach appears to be incorrect (see, e.g., Stein (2016)) and the future state of the market should be modelled using the real-world measure (\mathbb{Q} in \mathbb{P}). Real-world models are calibrated to the observed historical time-series and are typically used to compute non-traded quantities such as Value-at-Risk (VaR).

The risk-neutral measure at $t = 0$ is connected to the observed implied volatility surface and is therefore well-defined. However, the definition of the risk-neutral measure at future time $t = 1$ is debatable. Despite some relevant research on predicting the implied volatility surfaces (see, e.g., Cont et al. (2002), Mixon (2002) and Audrino and Colangelo (2010)), it is common practice to use option pricing models that are only calibrated at time $t = 0$, thereby assuming that the risk-neutral measure is independent of the state of the market (see, for example, Bauer et al. (2010) and Devineau and Loisel (2009)). This is however not in line with historical observations, where we see that the implied volatility surface does depend on the state of the market. Another drawback of this approach is that the resulting SCR is *pro-cyclical*, i.e., the SCR is relatively high when the market is in crisis and relatively low when the market is stable. The undesired effect of pro-cyclicality is that it can aggravate a downturn Bikker and Hu (2015).

In this paper, we investigate the impact of relaxing the assumption that the risk-neutral measure is considered to be independent of the state of the market and develop the so-called *VIX Heston model*, which depends on the current and also on simulated implied volatilities. This approach, which we have named here the $\mathbb{P}^{\mathbb{Q}}$ approach, takes into account the \mathbb{Q} measure information at time $t = 0$ and simulates risk-neutral model parameters (thus, future implied volatility surfaces are obtained by means of simulation) based on historically observed relations with some relevant market variables such as the VIX index.

As is well-known, the VIX index is a volatility measure for the S&P-500 index, which is calculated by the Chicago Board Options Exchange (CBOE) (see CBOE (2015)), and it is therefore directly linked to the implied volatility surface. Consequently, extracting information from the VIX index is frequently studied (see, e.g., Duan and Yeh (2012)) and our approach in this paper is based on the methodology presented in Singor et al. (2017), where the development of the Heston model parameters for the S&P-500 index options and the VIX index have been analysed.

The contribution of our present research is two-fold. First, we discuss the justification of using a risk-neutral model with time-dependent parameters. By means of a hedge test, we show that hedging strategies that take into account the changes in the implied volatility surface significantly outperform those strategies that do not, both in simulation and with empirical tests. This leads to the conclusion that the time-dependent risk-neutral measure can be used for the evaluation of future embedded option prices. Secondly, we show the impact of our new approach. For that, we use real data from 2007 to 2016 and compute the SCR on a monthly basis with a constant \mathbb{Q} measure and also with the VIX Heston model where this assumption is relaxed. We conclude that the VIX Heston model predicts out-of-sample implied volatility surfaces accurately and computes more conservative and stable SCRs. The impact of using the new approach on the SCR depends on the initial state of the market and may vary from -46% to $+52\%$ in our experiments. Moreover, we see that the SCR that is computed with the VIX Heston model is significantly less pro-cyclical, for example, it is lower in the wake of the 2008 credit crisis, as it incorporates the likely normalisation.

The outline of this paper is as follows. In Section 2, we give the definition of the SCR. In Section 3, we explain the dynamic VIX Heston model. In Section 4, we present the hedge tests with the corresponding results, followed by Section 5 where we present the numerical VIX Heston results and the impact of the using $\mathbb{P}^{\mathbb{Q}}$ dynamics on the SCR. Section 6 concludes.

2. Solvency Capital Requirement

Let us denote the policy’s net income in the interval $[0, t]$ by A_t , which is defined as the cash flows up to time t that are generated under the real-world market. The mathematical definition is given by

$$A_t = \int_0^t e^{\mu(t-s)} \text{cashflow}(s, \Delta s) ds, \tag{1}$$

where μ is the expected return and “cashflow($s, \Delta s$)” denotes the generated cash flow over the interval $[s, s + \Delta s]$. Similarly, we define the policy’s liabilities by L_t and they are given by the discounted expected cash flows under the risk-neutral measure in the interval $[t, T]$:

$$L_t = \mathbb{E}^{\mathbb{Q}_t} \left[\int_t^T e^{-r(s-t)} \text{cashflow}(s, \Delta s) ds \middle| \mathcal{F}_t \right], \tag{2}$$

with risk-free rate r . Note that a positive/negative cash flow corresponds to an income/liability for the insurer. We define

$$N_t := A_t - L_t, \tag{3}$$

which can be thought of as the policy’s net value at time t . The Solvency Capital Requirement is now defined as the 99.5% Value-at-Risk (i.e., the 99.5% quantile) of the one-year loss distribution under the real-world measure, i.e.,

$$\text{SCR} = \text{VaR}_{0.995} (N_0 - \tilde{N}_1) := \inf \{ x \mid \mathbb{P} (N_0 - \tilde{N}_1 < x) > 0.995 \}. \tag{4}$$

Here, \tilde{N}_1 is defined as the discounted value of N_1 .

Guaranteed Minimum Accumulation Benefit

This section is dedicated to deriving the fund dynamics and the SCR of a frequently used guarantee in the insurance industry, namely the Guaranteed Minimum Accumulation Benefit (GMAB) variable annuity rider.

We assume that the fund only contains stocks, but the derivation is similar when different assets are combined. We denote the stock and fund value by S_t and F_t , respectively, and define the initial premium by G . We assume the payout at maturity T is at least equal to the initial premium, in other words,

$$\text{Payout}_T = \max (F_T, G). \tag{5}$$

The dynamics of the fund are very similar to the stock dynamics, except for a fee α which is deducted from the fund as a payment to the insurer. This fee can be thought of as a dividend yield. Following Milevsky and Salisbury (2001), we obtain

$$dF_t = dS_t \frac{F_t}{S_t} e^{-\alpha t}, \quad F_0 = G. \tag{6}$$

The specific dynamics depend on the assumptions regarding stock price S_t . Here, we assume the Black–Scholes model (geometric Brownian motion, GBM) under the *observed real-world measure* \mathbb{P} , which yields:

$$dF_t^{\mathbb{P}} = (\mu - \alpha) F_t^{\mathbb{P}} dt + \sigma F_t^{\mathbb{P}} dW_t^{\mathbb{P}}, \quad F_0 = G. \tag{7}$$

Moreover, for the valuation, a risk-neutral Heston model is implemented, leading to

$$\begin{cases} dF_t = (r - \alpha) F_t dt + \sqrt{v_t} F_t dW_t^F, & F_0 = G, \\ dv_t = \kappa(\bar{v} - v_t) dt + \gamma \sqrt{v_t} dW_t^v, \\ \langle dW_t^S, dW_t^v \rangle = \rho dt. \end{cases} \tag{8}$$

The income is generated by the accumulated fees, hence

$$A_t = \int_0^t \alpha F_s^{\mathbb{P}} e^{\mu(t-s)} ds. \tag{9}$$

The liabilities, on the other hand, depend on the final value of the fund, as follows,

- If $F_T \geq G$, the policyholder receives F_T and the insurer has no liabilities.
- If $F_T < G$, the policyholder receives G and the liabilities of the insurer are equal to $G - F_T$.

Moreover, the insurer continues to claim future fees, hence, according to Equation (2), we can write the liabilities as

$$\begin{aligned} L_t &= \mathbb{E}^{\mathbb{Q}_t} \left[e^{-r(T-t)} \max(G - F_T, 0) - \int_t^T e^{-r(s-t)} \alpha F_s ds \middle| \mathcal{F}_t \right] \\ &= \text{Put}(F_t, G) + F_t \left(e^{-\alpha(T-t)} - 1 \right), \end{aligned} \tag{10}$$

where $\text{Put}(F_t, G)$ denotes the value of a European put option on the fund at time t with strike price G and dividend yield α . We can substitute these definitions into Equation (4) to obtain

$$\begin{aligned} \text{SCR} &= \text{VaR}_{0,995} (N_0 - e^{-r} N_1) \\ &= \text{VaR}_{0,995} (A_0 - L_0 - e^{-r} (A_1 - L_1)) \\ &= e^{-r} \text{VaR}_{0,995} (\text{Put}(F_1, G) - g(F_1, 1)) - \text{Put}(F_0, G) + g(F_0, 0), \end{aligned} \tag{11}$$

with

$$g(F_t, t) = \int_0^t \alpha F_s e^{\mu(t-s)} ds + F_t \left(1 - e^{-\alpha(T-t)} \right), \tag{12}$$

which can be thought of as the sum of the realized and expected fees. In this case, the SCR depends on the real-world distribution of F_1 , which determines $g(F_1, 1)$ and also influences the risk-neutral valuation of $\text{Put}(F_1, G)$.

Variable annuity riders require a risk-neutral valuation at (future) time $t = 1$, as the liabilities are, by definition, conditional expectations under the risk-neutral measure. In the case of a GMAB rider, there is an analytic expression available, but often there is no such expression for L_t . Hence, the evaluation of the conditional expectation typically requires an approximation. In that case, often the Least-Squares Monte Carlo algorithm is used to approximate the conditional expectations at $t = 1$. The experiments presented in this paper were also performed for the Guaranteed Minimum Withdrawal Benefit (GMWB) variable annuity rider. As the results and conclusions were very similar as for the GMAB, in this paper, we restrict ourselves to the GMAB variable annuity.

The distributions of $g(F_1, 1)$ and $\text{Put}(F_1, G)$ in Equation (12) can be obtained by means of a *Monte Carlo simulation*,

- First, the net policy value N_0 is determined, according to definition Equation (3).
- Thereafter, the fund value F_t along with other explanatory variables is simulated according to the real-world measure up to time $t = 1$.
- Subsequently, the values of $\text{Put}(F_1, G)$ and $g(F_1, 1)$ are evaluated for each trajectory. The value of $g(F_1, 1)$ can be obtained directly from the trajectory of F_t , however, $\text{Put}(F_1, G)$ requires a risk-neutral valuation for which the risk-neutral measure at time $t = 1$ is required.
- Finally, the simulated values are combined to construct the loss distribution. The SCR corresponds to the 99.5% quantile of this distribution.

Some more detail about the Monte Carlo simulation is provided in Appendix A.

3. Dynamic Stochastic Volatility Model

When valuing options, one typically wishes to calibrate a risk-neutral model according to the market’s expectations, which are quantified by the implied volatility surface. This implied volatility surface can be used to extract European option prices for a wide range of maturities and strikes. The market expectation (and so the volatility surface), is however unknown at $t = 1$ and therefore practitioners typically use the implied volatility surface at $t = 0$ to calibrate the risk-neutral model parameters. In this *standard \mathbb{Q} in \mathbb{P} approach*, these parameters are assumed to be constant over time, i.e., the risk-neutral measure is independent of the real-world measure. Note that this approach is also used to compute risk measures from the Basel accords, such as credit value adjustment (CVA), capital valuation adjustment (KVA) and potential future exposure (PFE) (see Kenyon et al. (2015); Jain et al. (2016); Ruiz (2014)). However, regarding the computation of CVA, this quantity is typically hedged, and so using only the market expectation at $t = 0$ is sufficient.

In the $\mathbb{P}^{\mathbb{Q}}$ approach, we relax the assumption of independence. The calibrated risk-neutral model parameters are related to the simulated real-world scenarios.

3.1. Heston Model

We assume the Heston model as a benchmark. The Heston stochastic volatility model (Heston 1993) assumes the volatility of the stock price process to be driven by a CIR model, i.e., under the \mathbb{Q} measure,

$$\begin{cases} dS_t = rS_t dt + \sqrt{v_t} S_t dW_t^1, & S(0) = S_0, \\ dv_t = \kappa(\bar{v} - v_t) + \gamma\sqrt{v_t} dW_t^2, & v(0) = v_0 \\ \langle dW_t^1, dW_t^2 \rangle = \rho dt. \end{cases} \quad (13)$$

where r denotes the risk-free rate, κ the speed of mean-reversion, \bar{v} the long-term variance, γ the volatility of variance and ρ the correlation between asset price and variance. The risk-free rate is assumed to be constant throughout this research. To calibrate these parameters according to the market’s expectations, one wishes to minimize the distance between the model’s and market’s implied volatilities. For the Heston model, consider the following search space for the parameters

$$\begin{aligned} \Omega^{\text{Search}} &= D_\kappa \times D_{v_0} \times D_{\bar{v}} \times D_\gamma \times D_\rho \\ &= [0, 10] \times [0, 1] \times [0, 1] \times [0, 2] \times [-1, 1], \end{aligned} \quad (14)$$

where D_p denotes the search domain for parameter p . Using this search space, one is able to find the calibrated parameters at time t by minimizing the sum of squared errors:

$$\Omega_t^{\text{Heston}} = \arg \min_{\Omega \in \Omega^{\text{Search}}} \left(\sum_K \sum_T \left(\sigma^{\text{Market}}(t, K, T) - \sigma^{\text{Heston}}(t, \Omega, K, T) \right)^2 \right). \quad (15)$$

The set of parameters $\Omega_t^{\text{Heston}} = \{\kappa, v_0, \bar{v}, \gamma, \rho\}$ minimizing this expression is considered risk-neutral and reflects the market’s expectations.

When the market is subject to changes, its expectations will change accordingly. Hence, the implied volatility surface will evolve dynamically over time. Consequently, the Heston parameters may change over time. Figure 1 shows the monthly evolution of the Heston parameters from January 2006 to February 2017.

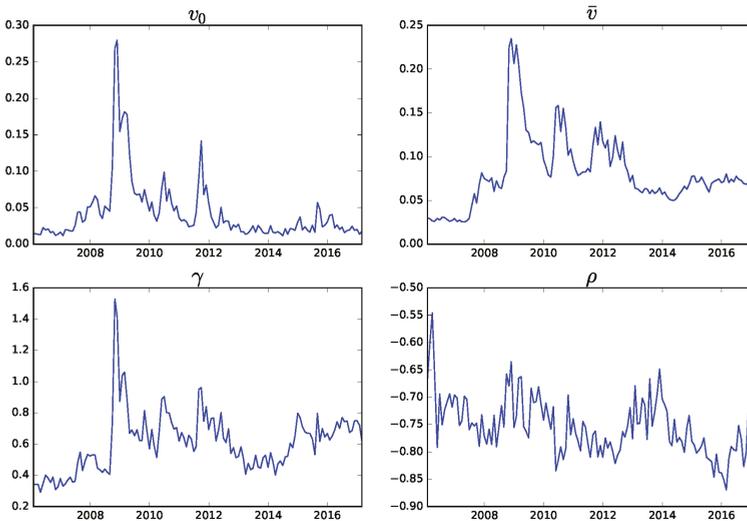


Figure 1. Evolution of the Heston parameters over time.

During this calibration procedure, we assumed κ to be constant; an unrestricted κ led to unstable results and did not significantly improve the accuracy. The other parameters, however, do not appear constant over time. This may give rise to issues in risk-management applications, where one simulates many real-world paths to assess the sensitivity to the market of a portfolio, balance sheet, etc. In many cases, a risk-neutral valuation is required, which is nested inside the real-world simulation, for example, when the portfolio or balance sheet contains options. Consequently, the future implied volatility surface for each trajectory needs to be known, such that the Heston parameters can be calibrated accordingly. Modelling the implied volatility surface over time is a challenging task (see, e.g., [Cont et al. \(2002\)](#); [Mixon \(2002\)](#); [Audrino and Colangelo \(2010\)](#)), as it quantifies the market’s expectations that depend on many factors. Moreover, even if the implied volatility surface is modelled, one would still need to perform a costly calibration procedure. Performing this calibration for each of the simulated trajectories would require a significant computational effort. As the Heston model is a parameterization of the implied volatility surface, an attractive alternative is to simulate the Heston parameters directly.

Figure 1 clearly shows that the parameters are time-dependent, which is in contrast with the assumptions of the plain Heston model. In this research we first search for relations between the risk-neutral parameters and observe real-world market indices, such as the VIX index. When a relation is found, the future risk-neutral parameters can be extracted from a real-world simulation. In this way, by performing a real-world simulation, we can directly forecast the set of Heston parameters within each simulated trajectory. The risk-neutral measure is then conditioned on the simulated state of the market, without the need of a simulated implied volatility surface.

3.2. VIX Heston Model

We have already described the difficulties of modelling the implied volatility surface, or equivalently, the option prices. A different approach is therefore required to calibrate the Heston parameters in simulated markets. The simulated parameter sets should accurately reflect the expectations of the simulated market, for example by linking the dynamics of the Heston parameters to the dynamics of the market. In [Singor et al. \(2017\)](#), an approach is considered which is based on the assumptions of a linear relationship between the VIX index and the Heston parameters. After analysis, it was concluded that:

- The initial volatility $\sqrt{v_{0,t}}$ and the volatility of the volatility γ_t are highly correlated to the VIX index, with correlation coefficients of 0.99 and 0.76, respectively.
- The long term volatility $\sqrt{\bar{v}_t}$ appears to be correlated to the VIX index trend line (estimated by a Kalman filter) with a correlation coefficient of 0.74.

To this end, the following restrictions are imposed on the Heston model parameters:

$$\Omega_t^{\text{Heston}}(X) = \begin{cases} \kappa_t &= \kappa, & \kappa \in \mathbb{R}_+, \\ v_{0,t} &= (a_{v_0} \cdot \text{VIX}_t + b_{v_0})^2, & a_{v_0}, b_{v_0} \in \mathbb{R}, \\ \bar{v}_t &= (a_{\bar{v}} \cdot \text{VIX}_{\text{filter}_t} + b_{\bar{v}})^2, & a_{\bar{v}}, b_{\bar{v}} \in \mathbb{R}, \\ \gamma_t &= a_{\gamma} \cdot \text{VIX}_t + b_{\gamma}, & a_{\gamma}, b_{\gamma} \in \mathbb{R}, \\ \rho_t &= \rho, & \rho \in [-1, 1], \end{cases} \quad (16)$$

where both the speed of mean reversion κ_t and the correlation coefficient ρ_t are assumed to be constant over time. The constant ρ assumption is justified by the fact that ρ displays a mean reverting pattern and it can therefore be approximated by its long-term mean. The constant κ assumption is justified by observations in Gauthier and Rivaille (2009). They argued that the effect on the implied volatility surface of increasing κ is similar to decreasing γ . Thus, allowing κ to change over time unnecessarily overcomplicates the model. Moreover, numerical experiments show that an unrestricted κ sometimes leads to unstable results.

The purpose of the restrictions is to accurately reflect the market’s expectations. To this end, we wish to minimize the distance between the observed and predicted implied volatility surfaces. Therefore, we calibrate the constraint parameters with a procedure similar to Equation (15). By changing the parameter set from Ω_t^{Heston} to $X = \{\kappa, a_{v_0}, b_{v_0}, a_{\bar{v}}, b_{\bar{v}}, a_{\gamma}, b_{\gamma}, \rho\}$ and summing over all points in time one obtains

$$X = \arg \min_{X^S \in X^{\text{Search}}} \left(\sum_{t,K,T} \left(\sigma^{\text{Market}}(t, K, T) - \sigma^{\text{Heston}}(t, \Omega_t^{\text{Heston}}(X^S), K, T) \right)^2 \right), \quad (17)$$

with,

$$\begin{aligned} X^{\text{Search}} &= D_{\kappa} \times D_{a_{v_0}} \times D_{b_{v_0}} \times D_{a_{\bar{v}}} \times D_{b_{\bar{v}}} \times D_{a_{\gamma}} \times D_{b_{\gamma}} \times D_{\rho} \\ &= \{1\} \times \mathbb{R}^6 \times [-1, 1]. \end{aligned} \quad (18)$$

By including the VIX-index in the real-world simulation, one is able to efficiently evaluate the set of Heston parameters in line with the simulated state of the market. For more information regarding the derivation and properties of the VIX Heston model, we refer the reader to Singor et al. (2017).

It is, however, important to stress the different assumptions in the real-world and risk-neutral markets. Risk-neutral valuations are performed under the Heston model, which assumes constant parameters. However, in real-world simulations, we assume the Heston parameters to *change over time*, according to the simulated state of the market, similar to Figure 1. One could argue that this approach is invalid, since we are violating the assumptions of the risk-neutral market. To this end, we discuss a justification of this approach, by means of a hedge test. Moreover, to assess the impact of time-dependent Heston parameters, we implement the VIX Heston model as proposed in Singor et al. (2017) in a risk-management application: the Solvency Capital Requirement.

4. Hedge Test

Before implementing the dynamic risk-neutral measure (the VIX Heston model) in a risk-management application, we first test its applicability from a theoretical point of view. For example, in theory, one should be able to hedge against future positions using today’s implied volatility surface. This no longer applies when one assumes a risk-neutral measure that changes over time. To this end,

we perform an experimental hedge test that determines which approach is more accurate in terms of future option prices, a dynamic or constant risk-neutral measure.

The plain Heston model assumes \bar{v} and γ to be constant, hence, from a theoretical point of view, it would be redundant to hedge against changes of these parameters. However, due to the dynamic behaviour of the implied volatility surface, \bar{v} and γ will change over time (see Figure 1). Thus, from an empirical point of view, the option price dynamics are subject to changes of these parameters. To support this claim, we compare three different hedging strategies. The first strategy is the *classical Delta-Vega hedge*, which does not take any changes of \bar{v} and γ into account. The replicating portfolio aims at hedging an option “A”, with value C_t , by holding a certain amount of stocks and a different option with value \bar{C}_t (which is called option “B” from this point onwards), i.e.,

$$\begin{cases} \Pi_t = -C_t + \Delta^{(1)}(t)S_t + \Delta^{(2)}(t)\bar{C}_t + B_t, \\ \Pi_0 = 0, \end{cases} \tag{19}$$

where B_t denotes the risk-free asset (for example, a bank account or a government bond), which grows with constant risk-free rate r . Note that Option B depends on the same underlying market factors as Option A. Following Bakshi et al. (1997), we impose the so-called minimized variance constraints,

$$\begin{cases} \langle d\Pi_t, dS_t \rangle = 0, \\ \langle d\Pi_t, dW_t^v \rangle = 0, \end{cases} \tag{20}$$

where $\langle \cdot, \cdot \rangle$ refers to the covariation between the two processes and W_t^v is the Brownian motion which is independent from S_t , driving random changes in the volatility. By imposing these constraints, one obtains a portfolio that has no covariation with the underlying asset and its volatility. In other words, changes in the asset’s value and changes in the asset’s volatility will have neither direct nor indirect (through correlations) effect on the portfolio. These constraints give us the following hedge ratios,

$$\begin{cases} \Delta^{(1)}(t) = \frac{\partial C}{\partial S_t} - \Delta^{(2)}(t) \frac{\partial \bar{C}}{\partial S_t}, \\ \Delta^{(2)}(t) = \frac{\partial C / \partial v_t}{\partial \bar{C} / \partial v_t}. \end{cases} \tag{21}$$

Under the assumptions of the Heston model, the portfolio dynamics are given by,

$$\begin{aligned} d\Pi_t = dt & \left(-\frac{\partial C}{\partial t} - \frac{1}{2}v_t S_t^2 \frac{\partial^2 C}{\partial S_t^2} - \frac{1}{2}\gamma^2 v_t \frac{\partial^2 C}{\partial v_t^2} - \rho\gamma v_t S_t \frac{\partial^2 C}{\partial S_t \partial v_t} + rB_t \right. \\ & \left. + \Delta^{(2)}(t) \left(\frac{\partial \bar{C}}{\partial t} + \frac{1}{2}v_t S_t^2 \frac{\partial^2 \bar{C}}{\partial S_t^2} + \frac{1}{2}\gamma^2 v_t \frac{\partial^2 \bar{C}}{\partial v_t^2} + \rho\gamma v_t S_t \frac{\partial^2 \bar{C}}{\partial S_t \partial v_t} \right) \right), \end{aligned} \tag{22}$$

Note that the random components have disappeared from the portfolio. Thus, the portfolio should be insensitive to changes in the market, if it respects the assumptions of the Heston model.

Secondly, we assume a model which is similar to the classical Delta-Vega hedge, but with adjusted hedge ratios. We call this strategy the *adjusted Delta-Vega hedge*. Assuming a dynamic model (see Appendix B for more details), we can apply Ito’s lemma to obtain

$$dC_t^{\text{Dynamic}} = \frac{\partial C}{\partial t} dt + \sum_{p_t} \frac{\partial C}{\partial p_t} dp_t + \frac{1}{2} \sum_{p_t, q_t} \frac{\partial^2 C}{\partial p_t \partial q_t} \langle dp_t, dq_t \rangle, \tag{23}$$

with $p_t, q_t \in \{S_t, v_t, \bar{v}_t, \gamma_t\}$. For notational purposes, we rewrite this expression as

$$dC_t^{\text{Dynamic}} = c_1 dt + c_2 dW_t^S + c_3 dW_t^v + c_4 dW_t^{\bar{v}} + c_5 dW_t^\gamma, \tag{24}$$

where the coefficients are defined as

$$\begin{cases} c_1 = \frac{\partial C}{\partial t} + rS_t \frac{\partial C}{\partial S_t} + \kappa(\bar{v}_t - v_t) \frac{\partial C}{\partial v_t} + \kappa_{\bar{v}}(\bar{v}_{\text{Mean}} - \bar{v}_t) \frac{\partial C}{\partial \bar{v}_t} \\ \quad + \kappa_{\gamma}(\gamma_{\text{Mean}} - \gamma_t) \frac{\partial C}{\partial \gamma_t} + \frac{1}{2} \sum p_i \sum q_i \frac{\partial^2 C}{\partial p_i \partial q_i} \langle dp_t, dq_t \rangle, \\ c_2 = \sqrt{\bar{v}_t} S_t \frac{\partial C}{\partial S_t} + \rho \gamma_t \sqrt{\bar{v}_t} \frac{\partial C}{\partial v_t} + \rho \rho_{\bar{v}} a_{\bar{v}} \bar{v}_t \frac{\partial C}{\partial \bar{v}_t} + \rho \rho_{\gamma} a_{\gamma} \gamma_t \frac{\partial C}{\partial \gamma_t}, \\ c_3 = \gamma_t \sqrt{v_t(1 - \rho^2)} \frac{\partial C}{\partial v_t} + \rho_{\bar{v}} a_{\bar{v}} \bar{v}_t \sqrt{1 - \rho^2} \frac{\partial C}{\partial \bar{v}_t} + \rho_{\gamma} a_{\gamma} \gamma_t \sqrt{1 - \rho^2} \frac{\partial C}{\partial \gamma_t}, \\ c_4 = a_{\bar{v}} \bar{v}_t \sqrt{1 - \rho_{\bar{v}}^2} \frac{\partial C}{\partial \bar{v}_t}, \\ c_5 = a_{\gamma} \gamma_t \sqrt{1 - \rho_{\gamma}^2} \frac{\partial C}{\partial \gamma_t}. \end{cases} \tag{25}$$

The hedge ratios now take the correlated components of \bar{v} and γ into account, due to the minimized variance constraints of Equation (20). Using Equation (24), the hedge ratios can be derived, giving

$$\begin{cases} \Delta_{\text{Adjusted}}^{(1)}(t) = \frac{c_2}{\sqrt{\bar{v}_t} S_t} - \frac{\tilde{c}_2 c_3}{\sqrt{\bar{v}_t} S_t \tilde{c}_3}, \\ \Delta_{\text{Adjusted}}^{(2)}(t) = \frac{c_3}{\tilde{c}_3}, \end{cases} \tag{26}$$

where \tilde{c}_2 and \tilde{c}_3 are defined as in Equation (25) for Option B. The additional stochastic variables \bar{v}_t and γ_t follow mean reverting processes (see Equation (A1) in Appendix B), such that the portfolio dynamics are found to be,

$$\begin{aligned} d\Pi_t = & \left(-c_1 + \frac{c_3}{\tilde{c}_3} \tilde{c}_1 + \frac{\tilde{c}_2 r}{\sqrt{\bar{v}_t}} - \frac{\tilde{c}_2 c_3 r}{\sqrt{\bar{v}_t} \tilde{c}_3} + rB_t \right) dt \\ & + \left(\frac{c_3}{\tilde{c}_3} \tilde{c}_4 - c_4 \right) dW_t^{\bar{v}} + \left(\frac{c_3}{\tilde{c}_3} \tilde{c}_5 - c_5 \right) dW_t^{\gamma}. \end{aligned} \tag{27}$$

The portfolio still depends on the randomness associated with \bar{v} and γ . However, the randomness associated with S_t and v_t have disappeared, including the random components of \bar{v} and γ that are correlated to S_t and v_t .

We also consider a strategy that aims at completely hedging against any changes of \bar{v} and γ , by introducing two additional options,

$$\begin{cases} \Pi_t = -C_t + \Delta_{\text{Full}}^{(1)}(t) S_t + \Delta_{\text{Full}}^{(2)}(t) \tilde{C}_t + \Delta_{\text{Full}}^{(3)}(t) \hat{C}_t + \Delta_{\text{Full}}^{(4)}(t) \hat{\hat{C}}_t. \\ \Pi_0 = 0. \end{cases} \tag{28}$$

Again, all options depend on the same underlying market factors, but they have different contract details. In this case, we require to be protected against any changes of S_t , v_t , \bar{v}_t and γ_t , hence we impose

$$\begin{cases} \langle d\Pi_t, dS_t \rangle = 0, \\ \langle d\Pi_t, dW_t^v \rangle = 0, \\ \langle d\Pi_t, dW_t^{\bar{v}} \rangle = 0, \\ \langle d\Pi_t, dW_t^{\gamma} \rangle = 0. \end{cases} \tag{29}$$

Substituting these constraints leads to a system of equations, which is solved by

$$\begin{bmatrix} \Delta_{\text{Full}}^{(1)}(t) \\ \Delta_{\text{Full}}^{(2)}(t) \\ \Delta_{\text{Full}}^{(3)}(t) \\ \Delta_{\text{Full}}^{(4)}(t) \end{bmatrix} = \begin{bmatrix} \sqrt{\bar{v}_t} S_t & \tilde{c}_2 & \tilde{c}_2 & \hat{c}_2 \\ 0 & \tilde{c}_3 & \tilde{c}_3 & \hat{c}_3 \\ 0 & \tilde{c}_4 & \tilde{c}_4 & \hat{c}_4 \\ 0 & \tilde{c}_5 & \tilde{c}_5 & \hat{c}_5 \end{bmatrix}^{-1} \begin{bmatrix} c_2 \\ c_3 \\ c_4 \\ c_5 \end{bmatrix}. \tag{30}$$

By imposing Equation (29), one removes all randomness associated with S_t , v_t , \bar{v}_t and γ_t . Hence, the portfolio dynamics only depend on deterministic changes under the assumed market dynamics. From this point onwards, we refer to this strategy as the *full hedge* strategy.

4.1. Hedge Test Experiments

In this section, we discuss the results of the hedge test which was explained above. This test experiment indicates which approach is more accurate when evaluating future option prices, a measure which assumes \bar{v} and γ to be constant or a measure which assumes \bar{v} and γ to change over time. We thus consider the classical approach, with constant \bar{v} and γ , and two approaches that assume \bar{v} and γ to change over time. The different hedging strategies are tested in a simulated market as well as in an empirical market.

The empirical hedge test is based on daily implied volatility surfaces of the S&P-500 European put and call options from January 2006 to February 2014.

4.1.1. Simulated Market

First, we evaluate the hedging strategies in a controlled environment. In this test, we assume S_t and v_t to follow the Heston model, with the key difference that \bar{v} and γ are time-dependent mean reverting processes (as in Equation (A1) in Appendix B). This set-up will be informative, as it is in line with historical observations. The processes of S_t and v_t are discretized by the Quadratic Exponential (QE) scheme, as proposed in Andersen (2008). Furthermore, the processes of \bar{v} and γ are discretized by the Milstein scheme, while also taking the correlations with v_t into account. The details of this simulation scheme can be found in Appendix B.

The performance of the classical Heston Delta-Vega, the adjusted Heston Delta-Vega and the full hedge is compared under the dynamics of this market. The strategies all aim at hedging a short position in a European call option with maturity $T = 1.0$ years and strike $K = 50$. Moreover, at the start date of the option, $t = 0$, we assume

$$\begin{cases} S_0 = 49, & v_0 = 0.05, & \bar{v}_0 = 0.1, & \gamma_0 = 0.7, \\ \kappa = 1.0, & \rho = -0.75, & r = 0.01. \end{cases} \quad (31)$$

The adjusted Delta-Vega and full hedge involve additional options. These options depend on the same stock as Option A, but the contract details are different, i.e.,

$$\begin{cases} \bar{K} = 50.0, & \bar{T} = 2.0, \\ \bar{K} = 50.0, & \bar{T} = 3.0, \\ \hat{K} = 50.0, & \hat{T} = 4.0. \end{cases} \quad (32)$$

Moreover, we assume the following parameters in the dynamics of the \bar{v} and γ processes,

$$\begin{cases} \kappa_{\bar{v}} = 1.4, & \bar{v}_{\text{Mean}} = 0.1, & a_{\bar{v}} = 0.8, & \rho_{\bar{v}} = 0.9, \\ \kappa_{\gamma} = 2.1, & \gamma_{\text{Mean}} = 0.7, & a_{\gamma} = 1.0, & \rho_{\gamma} = 0.9. \end{cases} \quad (33)$$

In reality, these parameters cannot be freely chosen, as they are implied by the market. By analysing the historical behaviour of \bar{v}_t and γ_t , one is able to estimate these SDE parameters. In this case, however, we assume to know these parameters and use them when determining the hedge strategy.

Ideally, the portfolio value should be equal to zero for each point in time, because the initial portfolio value is equal to zero. Every deviation from zero is thought of as a hedge error. Over the entire life of the option we desire the mean and standard deviation of this error to be as close as possible to zero. To this end, we introduce the following error measures for simulation j ,

$$E_{\text{Mean}}^{(j)} = \frac{1}{N+1} \sum_{i=0}^N \Pi_{\frac{i}{N}T}^{(j)}, \quad E_{\text{Std}}^{(j)} = \sqrt{\frac{1}{N} \sum_{i=0}^N \left(\Pi_{\frac{i}{N}T}^{(j)} - E_{\text{Mean}}^{(j)} \right)^2}. \tag{34}$$

The overall hedging performance of the M simulations can be judged by

$$\begin{cases} \bar{E}_{\text{Mean}} = \frac{1}{M} \sum_{j=1}^M E_{\text{Mean}}^{(j)}, \\ \bar{E}_{\text{Std}} = \frac{1}{M} \sum_{j=1}^M E_{\text{Std}}^{(j)}. \end{cases} \tag{35}$$

In the current set-up, these error measures are random variables, as they are determined by means of Monte Carlo simulations. To this end, we analyse the stability of the error measures across the simulated trajectories by the standard error,

$$\text{SE}(E) = \frac{\sqrt{\frac{1}{M-1} \sum_{j=1}^M (\bar{E} - E^{(j)})^2}}{\sqrt{M}}, \tag{36}$$

with error measure E , its mean \bar{E} and the simulated trajectories $E^{(j)}$.

The performance of the strategies under these assumptions based on $M = 200$ simulations is given in Table 1.

Table 1. Hedge errors of the different hedging strategies. The standard errors of the estimates are given in parentheses.

Frequency	\bar{E}_{Mean}	\bar{E}_{Std}
<i>classical Delta-Vega:</i>		
Once per week	-0.339 (0.0207)	0.228 (0.0084)
Once per day	-0.354 (0.0227)	0.210 (0.0080)
<i>Adjusted Delta-Vega:</i>		
Once per week	-0.204 (0.0097)	0.137 (0.0072)
Once per day	-0.220 (0.0071)	0.114 (0.0047)
<i>Full hedge:</i>		
Once per week	-0.044 (0.0040)	0.068 (0.0042)
Once per day	-0.051 (0.0030)	0.045 (0.0022)

These results show that, in this experiment, it is beneficial to take parameter correlations into account when hedging, in terms of both the mean error and the standard deviation. While still not perfect, the dynamic Heston Delta-Vega hedge is better able to remain risk-neutral on average and it deviates less from this average. The full hedge performs even better with a mean approximately equal to zero and a standard deviation equal to or lower than any of the previous strategies, despite the dynamic behaviour of the market.

The purpose of these hedging strategies is to replicate the value of an option. In the case of the classical Delta-Vega hedge, only S_t and v_t are allowed to change. By respecting the assumptions of the Heston model, we are not fully able to replicate the future option values. On the other hand, when assuming the adjusted Heston model, \bar{v}_t and γ_t are allowed to change as well. Hedging strategies considering this dynamic behaviour, produce more accurate future option price estimates in the current set-up. This indicates the importance of taking dynamic parameters into account when determining future option prices, even when the assumptions of the underlying model are violated. However, note that the comparison is not completely fair, since we specifically assume a Heston market with time-dependent \bar{v} and γ . It can therefore be expected that a strategy taking these assumptions into account outperforms one that does not. Therefore, to better assess the *true* performance of these strategies, we also perform an empirical test.

4.1.2. Empirical Market

When hedging in practice, the underlying assumptions are not always respected and the *true* parameters are, of course, unknown. To quantify the effect of these difficulties, we test the hedging strategies on historical data in this section. All hedging ratios depend on the dynamic risk-neutral parameters, hence v_t , \bar{v} , γ and ρ vary over time, according to the changes in the historically observed implied volatility surfaces. Moreover, the adjusted Heston Delta-Vega hedge depends on the correlation and volatility of \bar{v} and γ , which we define as follows

$$\begin{cases} \rho_{\bar{v}} = \rho_{\gamma} = 0.95, \\ a_{\bar{v}} = \sqrt{\frac{\frac{1}{N} \sum_{i=-N}^{-1} \log\left(\frac{\bar{v}_{t_{i+1}}}{\bar{v}_{t_i}}\right)^2}{\Delta t}}, \\ a_{\gamma} = \sqrt{\frac{\frac{1}{N} \sum_{i=-N}^{-1} \log\left(\frac{\gamma_{t_{i+1}}}{\gamma_{t_i}}\right)^2}{\Delta t}}. \end{cases} \tag{37}$$

It can be quite challenging to determine the correlation between the Brownian motions driving the parameters, hence we assume it to be equal for each time-interval. Moreover, the volatility estimator only depends on past observations; the sum indices vary from $-N$ to -1 with $t_{-N} = -1$ year, where $t_0 = 0$ indicates the starting date of the option.

In this test, we hedge an at-the-money European call option with one year maturity on the S&P-500 index. All hedging strategies are subjected to daily rebalances that are based on the parameters as seen on that date. The transaction costs are excluded from this test, as we are interested in the performance of the hedging strategies with respect to changes in \bar{v} and γ . Including transaction costs would increase the costs of the full dynamic Heston hedging strategy, as it involves more financial assets. This would bias the results and it is therefore best to exclude the transaction costs from the present test.

The test is repeated on a monthly basis from July 2006 to February 2013 and the performance is assessed by the mean error and mean squared error during the life of the option,

$$E_{\text{Mean}}^{(j)} = \frac{1}{N+1} \sum_{i=0}^N \Pi_{\frac{i}{N}T}^{(j)}, \quad E_{\text{MSE}}^{(j)} = \frac{1}{N+1} \sum_{i=0}^N \left(\Pi_{\frac{i}{N}T}^{(j)} \right)^2. \tag{38}$$

The time intervals of the hedging portfolios overlap in this set-up, since the test is repeated on a monthly basis and the option maturity is one year. However, all strategies depend on different initial conditions and therefore perform differently, despite the overlapping time-intervals. The results of this test are graphically presented in Figure 2.

The hedging performances are similar to the simulation results:

- The classical Delta-Vega hedge does not take changes of the parameters into account and appears to be the most unstable method. This strategy has the most and highest error “peaks” and is therefore most unreliable.
- The adjusted Delta-Vega hedge is still not perfect but appears to be more stable than the classical Delta-Vega hedge, the error “peaks” happen less frequently and are less pronounced. The error can be minimized by optimizing the correlation and volatility of \bar{v}_t and γ_t , but the optimization can only be performed afterwards, which is not the objective of this test.
- The full hedge is the most stable out of the three strategies. It does not have any error “peaks” and outperforms the other two strategies in most cases. Moreover, this strategy does not depend on additional parameters which may introduce an error if chosen poorly, such as in the dynamic Heston Delta-Vega hedge.

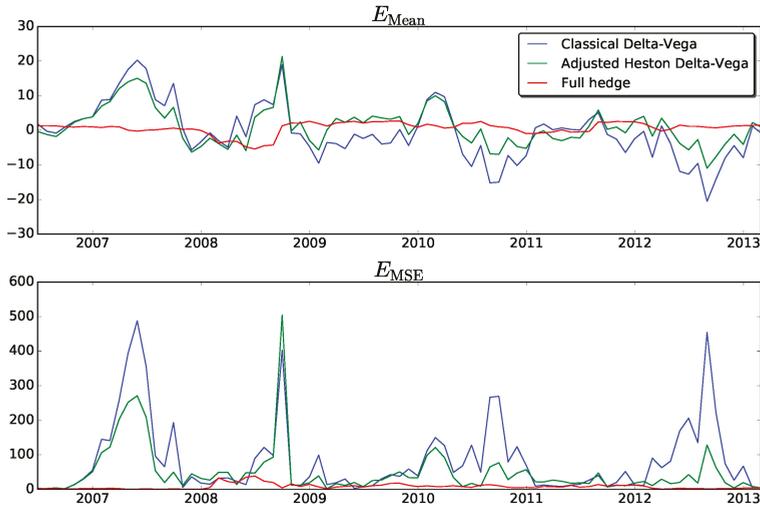


Figure 2. Mean error and mean squared error for different hedging strategies performed on monthly historical data.

We can conclude that respecting the assumptions of the underlying model (in this case, the Heston model) does not necessarily lead to more accurate future option prices. By taking changes of the \bar{v} and γ parameters into account, we are able to replicate option values more accurately both in a controlled (simulation) and uncontrolled (empirical) environment.

5. VIX Heston Model Results

5.1. Data and Calibration

The dataset contains monthly implied volatility surfaces of the S&P-500 European put and call options from January 2006 to February 2017. Each implied volatility surface contains five different strike levels (80%, 90%, 100%, 110% and 120% of S_0) and maturities (0.25, 0.5, 1, 1.5 and 2 years). We split the dataset into a training set and a test set. The training set consists of implied volatility surfaces from January 2006 to February 2014 and is only used to identify the optimal regression components. The test set contains monthly implied volatility surfaces from March 2014 to February 2017 and is used to assess the accuracy of the VIX Heston model.

Furthermore, to assess the robustness of the VIX Heston model, we apply it to monthly implied volatility surfaces of the FTSE-100 (United Kingdom) and STOXX-50 (Europe) as well. The training set includes data from October 2010 to June 2015 and the test set contains data from July 2015 to February 2017.

To assess the accuracy of the VIX Heston model, one must first calibrate the model according to Equation (17). Using the US training set described above, we obtain

$$\Omega_t^{\text{Heston}}(X) = \begin{cases} \kappa_t & = 1.0, \\ v_{0,t} & = (0.0140 + 0.0090 \cdot \text{VIX}_t)^2, \\ \bar{v}_t & = (0.0957 + 0.0087 \cdot \text{VIX}_{\text{filter}_t})^2, \\ \gamma_t & = 9.6479 \cdot 10^{-5} + 0.0270 \cdot \text{VIX}_t, \\ \rho_t & = -0.7294. \end{cases} \quad (39)$$

The calibrated parameters of the UK and Europe datasets can be found in Appendix C. The accuracy is assessed by comparing the predicted to the observed implied volatility surfaces of the test set, according to the following error measures,

$$\left\{ \begin{array}{l} \text{SSE} = \sum_{t,K,T} \left(\sigma^{\text{Market}}(t,K,T) - \sigma^{\text{Heston}}(t, \Omega_t^{\text{Heston}}, K, T) \right)^2, \\ \text{MAE} = \frac{1}{N_\sigma} \sum_{t,K,T} \left| \sigma^{\text{Market}}(t,K,T) - \sigma^{\text{Heston}}(t, \Omega_t^{\text{Heston}}, K, T) \right|, \\ R^2 = 1 - \frac{\text{SSE}}{\sum_{t,K,T} \left(\sigma^{\text{Market}}(t,K,T) - \bar{\sigma}^{\text{Market}} \right)^2}, \\ R_{\text{Min}}^2 = \min_t \{ R_t^2 : t \in [t_{\text{min}}, t_{\text{max}}] \}, \end{array} \right. \quad (40)$$

where Ω_t^{Heston} is defined as the predicted parameter set, N_σ as the total number of observed implied volatilities and $\bar{\sigma}^{\text{Market}}$ as the average of all observed implied volatilities. The corresponding results are displayed in Table 2. The predicted paths for the Heston parameters of the US dataset can be found in Figure 3. The predicted paths of the Heston parameters of the UK and Europe dataset are graphically presented in Appendix C.

Table 2. Out-of-sample accuracy of the regression models according to the error measures defined in Equation (40).

	SSE	MAE	R ²	R ² _{Min}
<i>US</i>				
VIX Heston	0.2286	0.0124	0.8948	0.8159
Unrestricted	0.0185	0.0034	0.9915	0.9798
<i>UK</i>				
VIX Heston	0.0577	0.0093	0.9340	0.7401
Unrestricted	0.0126	0.0045	0.9855	0.9690
<i>Europe</i>				
VIX Heston	0.0433	0.0078	0.9389	0.7693
Unrestricted	0.0154	0.0051	0.9783	0.9190

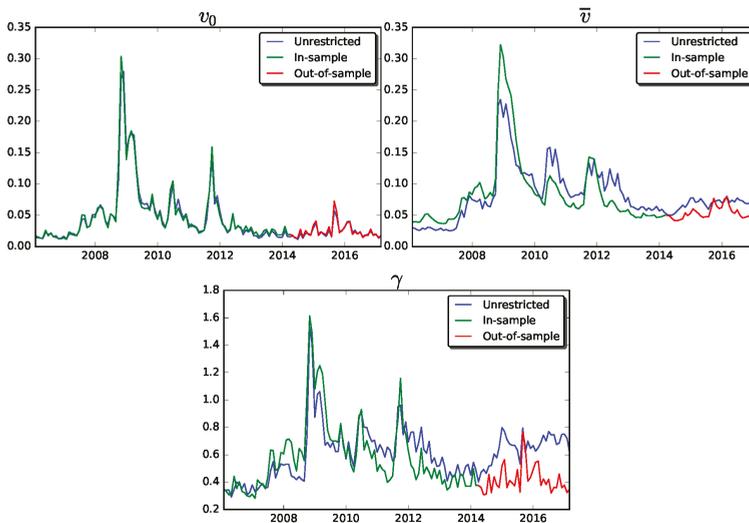


Figure 3. Prediction results Heston parameters of the US dataset.

The regression model loses some accuracy compared to the unrestricted model. On average, there is an error of 0.003 between the implied volatility and the unrestricted Heston model in the US dataset, which is approximately equal to an error of 2.2%. The VIX Heston model has an average absolute error of 0.012, which corresponds to a 7.7% error. Thus, by implementing the regression models, we introduce an additional error of 5.5%, on average. The accuracy of the VIX Heston model in the UK and Europe datasets is even higher, where the accuracy loss is equal to 2.9% and 1.2%, respectively.

Finally, we discuss the predictions obtained in dependence of parameter γ . In the US dataset, the prediction of γ is relatively inaccurate (see Figure 3), as the out-of-sample correlation to the VIX index is much lower than the in-sample correlation (0.86 in-sample versus 0.32 out-of-sample). In the UK and Europe datasets, this phenomenon does not seem to be present and consequently the predictions of γ are much more accurate (see Figures A1 and A2 in Appendix C). This also explains why the implied volatility surface predictions in the UK and Europe datasets are more accurate than the US predictions, as can be seen in Table 2. Thus, there appears to be another, yet unknown, factor driving γ in the US dataset, which is absent in the UK and Europe datasets. Analysis of the cause of this phenomenon might be a topic for future study. However, in this research, we assume that the VIX Heston model is sufficiently accurate to describe the dynamic behaviour of the Heston parameters, as γ only has a minor effect on the implied volatility surface.

5.2. SCR Impact Study

In this test, we assess the impact of assuming time-dependent risk-neutral parameters, according to three scenarios. Each scenario corresponds to a different initial market and consequently the initial expected liabilities, L_0 , will differ. We assume a no-arbitrage fee, i.e., a fee for which L_0 is equal to zero. The general contract details of the variable annuity can be found in Table 3 and the initial values of the different scenarios accompanied with the fair premiums are presented in Table 4.

Table 3. General contract details of the GMAB rider.

Parameter	Value
F_0	1000
G	1000
T_{GMAB}	10
r	0.04

Table 4. Initial values and fair premiums of the GMAB rider.

Parameter	Scenario 1	Scenario 2	Scenario 3
v_0	0.04	0.01	0.27
\bar{v}_0	0.08	0.025	0.24
γ_0	0.55	0.05	1.4
α_{GMAB}	0.0174	0.0057	0.0345

We assume the fund value to follow a GBM with initial value $F_0 = 1000$, so the fund value follows Equation (7). The VIX index is modelled simultaneously, following a mean-reverting path

$$\begin{cases} dvix_t &= \kappa_{vix}(vix_{Mean} - vix_t) + \gamma_{vix}vix_t^{\lambda_{vix}} \left(\rho_{vix}dW_t^S + \sqrt{1 - \rho_{vix}^2}dW_t^{vix} \right), \\ VIX &= 100 \cdot vix. \end{cases} \tag{41}$$

The process parameters are estimated with the generalized Method of Moments (see Hansen (1982)),

$$\begin{cases} \sigma = 0.21, & \mu = 0.05, & \kappa_{vix} = 4.964, & vix_{Mean} = 0.207, \\ \gamma_{vix} = 1.859, & \lambda_{vix} = 1.271. \end{cases} \tag{42}$$

Moreover, we set $\rho_{vix} = -0.75$, which is in line with observations in the risk-neutral market. The SDEs are discretized by the Milstein scheme.

5.2.1. Guaranteed Minimum Accumulation Benefit

We have simulated 100,000 real-world trajectories and evaluated the loss function as defined in Equation (4) under the two risk-neutral measures (with either constant or time-dependent parameters). This way, we are able to construct and compare the probability density functions under these measures. This process is repeated for the three different scenarios that are presented in Table 3. In Figure 4, we have graphically represented the impact on the probability density function of the loss distribution for the different scenarios. Moreover, the Solvency Capital Requirements associated with these distributions can be found in Table 5.

Table 5. Solvency Capital Requirement of the scenarios for the original and time-dependent risk-neutral measure.

	Original	Time-Dependent
Scenario 1	170.1	195.5
Scenario 2	166.9	237.5
Scenario 3	178.8	150.8

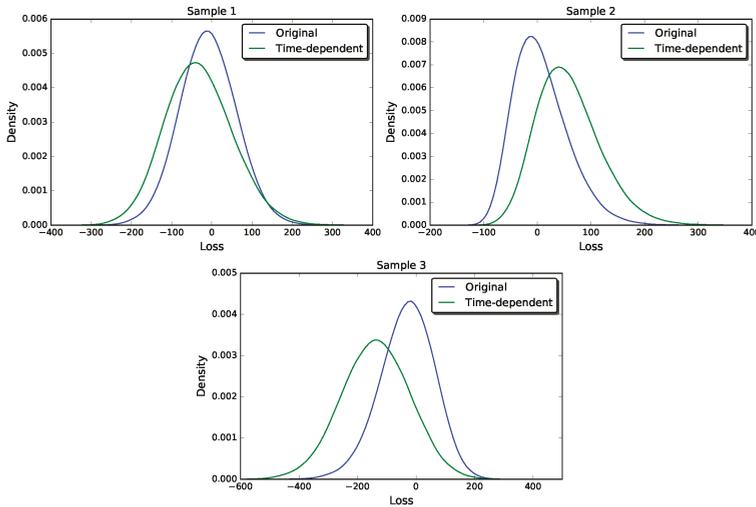


Figure 4. Probability density functions of the one-year loss distribution for a variable annuity with the GMAB rider, under the original and time-dependent risk-neutral measure. Scenario 1 = average initial volatility; Scenario 2 = low initial volatility; and Scenario 3 = high initial volatility.

The impact on the probability density functions and the Solvency Capital Requirements is substantial and we wish to highlight a few noteworthy features.

The loss distribution under the original risk-neutral measure appears to be centred around 0, independent of the initial conditions. The one-year loss is defined as the difference between the policy value at the times $t = 0$ and $t = 1$. On average, the policy value will not change significantly if the risk-neutral parameters stay the same. Therefore, the loss distribution must be centred around 0, as long as the initial risk-neutral parameters do not change. The loss distribution under the time-dependent risk-neutral measure, on the other hand, heavily depends on the risk-neutral parameters at $t = 0$. Consider Scenario 2 for example. Initially, the volatility, \bar{v} and γ are relatively low, resulting in low

initial expected liabilities. However, according to the mean-reverting VIX index, these parameters are more likely to increase over time and with them, the expected liabilities. Consequently, the one-year losses are much higher compared to those under the original risk-neutral measure, which still assumes the relatively low initial parameters at $t = 1$. In Figure 4, this effect is clearly visible where the loss function is shifted to the right. Conversely, the one-year losses under the time-dependent risk-neutral measure in Scenario 3 are much lower, as the expected liabilities are more likely to decrease. In conclusion, when the initial volatility is low (high), we can expect a higher (lower) SCR under the time-dependent risk-neutral measure. In the 2008 credit crisis, this resulted in a higher SCR before the crisis and a lower SCR during the crisis.

Besides the shifted mean, the loss distribution under the time-dependent risk-neutral measure also tends to have heavier tails, which is especially visible in Scenario 1. This is caused by the fact that \bar{v} and γ depend on the state of the market, which results in more extreme losses (or gains). If, for example, the market crashes, \bar{v} and γ are likely to increase. This will generate even higher expected liabilities, resulting in even higher losses. However, if the market flourishes, \bar{v} and γ tend to be much lower, leading to lower expected liabilities and lower losses (or higher gains). This feature is present in all scenarios of Figure 4, but is best visible in Scenario 1, where the probability of an extreme loss as well as the probability of an extreme gain is higher under the time-dependent risk-neutral measure. Consequently, the SCRs under the two risk-neutral measures are not necessarily equal, not even when the initial conditions are equal to the average market conditions (such as in Scenario 1).

To give a broad overview of the impact, we determine the SCR of a variable annuity with the GMAB rider for multiple points in time. The contract details presented in Table 3 remain unchanged, but the initial parameters depend on historical data. For computational purposes, the parameter α is assumed to be constant and equal to 0.01. In this test, we compare four different risk-neutral measures:

1. A time-dependent risk-neutral measure where all parameters depend on the simulated state of the market.
2. A risk-neutral measure where F_1 and v_1 depend on the simulated market and the risk-neutral parameters are equal to the parameters as observed on $t = 0$. This measure is equivalent to the original risk-neutral measure that we have previously defined.
3. A risk-neutral measure where F_1 and v_1 depend on the simulated market and the risk-neutral parameters are equal to the parameters as observed on $t = 1$. We refer to this measure as the future risk-neutral measure.
4. A risk-neutral measure where F_1 and v_1 depend on the simulated market and the risk-neutral parameters are equal to the *realized* regression model predictions at $t = 1$ of Figure 3. This measure is different from the time-dependent risk-neutral measure, as it depends on the realized state VIX, instead of the simulated VIX. We refer to this measure as the future VIX risk-neutral measure.

Thus far, we have applied the first two risk-neutral measures in our analysis. The latter two measures can only be applied on historical data (otherwise, the observed parameters at $t = 1$ are undefined) and are merely added for explanatory purposes. Ideally, the SCRs under the future and the future VIX risk-neutral measure are equal. The difference between these measures is caused by prediction errors of the regression model. Hence, the difference between the SCRs under these measures be is indication for the accuracy of the regression models.

In Figure 5, the results under the different risk-neutral measures are displayed. The difference between the original and time-dependent risk-neutral measure is also summarized in Table 6.

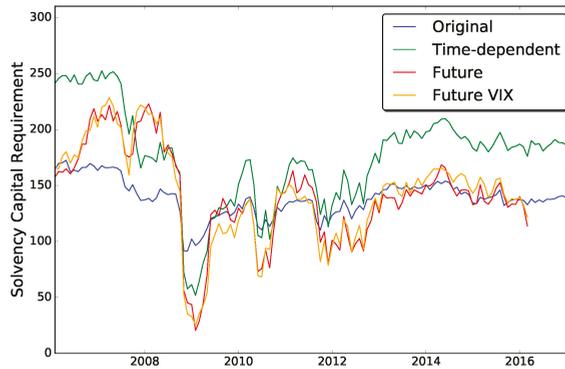


Figure 5. SCR over time under the different risk-neutral measures.

Table 6. Difference in SCR between original and time-dependent risk-neutral measure.

	Value	%
Mean absolute difference	41.1	28.7%
Maximum absolute difference	85.3	52.0%

5.2.2. Discussion on Impact

The impact on the SCR is significant, with a maximum absolute difference of over 50%. Moreover, the difference appears to be structural over time, with a mean absolute difference of almost 30%. The results in Figure 5 also contain some stylized facts that we have already seen in Figure 4 and we discuss them by distinguishing different time periods:

- 2005–2007: The volatility in these years was relatively low and this translates to a somewhat higher SCR under the original measure and an even higher SCR under the time-dependent measure, which is similar to Scenario 2. The SCR under the future measure is rising, due to higher expected liabilities at $t = 1$, indicating that more volatile times are coming.
- Early 2008: The market has not crashed yet, but volatility is starting to increase, resulting in a smaller difference between the time-dependent and original measures, which is comparable to Scenario 1. The future measure, however, takes the fact that the market will crash into account. Hence, the SCR is the highest under the future measure.
- Late 2008–2012: During these years, several spikes occurred in the implied volatility surface, which will increase the initial liabilities and therefore the expected losses will decrease, analogously to Scenario 3. This results in lower SCRs during this period. Moreover, we see that the SCRs under the future risk-neutral measure are lowest during these highly volatile periods, since this measure depends on the realized market at $t = 1$, which has returned to its less volatile state. Consequently, the expected liabilities at $t = 1$ and the SCR are the lowest under the future risk-neutral measure.
- 2013–2017: This period is comparable to 2005–2007, apart from the fact that the volatility is approximately constant throughout these years. This translates into an almost equal SCR prediction under the original and future risk-neutral measures.

The difference between the future and future VIX measures is small. This means that the expected liabilities at $t = 1$ are almost equal under both measures, indicating the accuracy of the regression model, at least for the realized states of the market. Under the assumption that the simulated markets behave similar to historical observations, this means that expected liabilities at $t = 1$ under the time-dependent measure will be in line with the simulated states of the market.

Finally, we also mention that we performed a very similar study for the GMWB product. Essentially, the same results were obtained in that case, however, the impact was somewhat less severe.

6. Conclusions

The research in this paper was motivated by the open question of how to value future guarantees that are issued by insurance companies. The future value of these guarantees is essential for regulatory and Asset Liability Management purposes. The complexity of the valuation is found in the fact that, first, these guarantees involve optionalities and thus need to be valued using the risk-neutral measure; and, second, whereas this measure is well-defined at $t = 0$, the future risk-neutral measure, at future time $t = 1$, is debatable.

For a large part, the liabilities evolve according to real-world models and, therefore, the future values of these guarantees need to be computed conditionally on the real-world scenarios. In this paper, we demonstrate the benefits of option valuation under a new, so-called \mathbb{P}^Q measure in Asset Liability Management. This is done by modelling the Heston model parameters, which form the parameterization of the implied volatility surface, conditional on the real-world scenarios.

Basically, we advocate the use of dynamic risk-neutral parameters in the cases in which we need to evaluate asset prices under the \mathbb{P} measure, before an option value is required at a future time point. It means that the development of the real-world asset paths in the future are taken into account in the option valuation.

A hedge test was implemented for an academic test case, where the dynamic strategy outperformed the strategy with static parameters. Importantly, the results from this hedge test case were confirmed by a hedge test based on 12 years of empirical, historical data. Several conclusions have already been given after each of the structured test experiments is presented.

The results obtained by the strategy for the Solvency Capital Requirement of the variable annuities exhibited differences of even 50%, as compared to the conventional risk-neutral pricing of these annuities. Next to that, we saw that the SCR was significantly less pro-cyclical under the new approach, which is a highly desired feature.

Author Contributions: M.T.P.D., Data Curation; C.S.L.G., Formal Analysis; C.W.O., Methodology.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank Pieter Kloek for helpful discussions on option valuation in the context of the SCR.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Least-Squares Monte Carlo Method

In this section, we briefly describe the numerical techniques employed, which are based on the well-known least-squares Monte Carlo method.

The Least-Squares Monte Carlo method was first proposed by Longstaff and Schwartz (2001) for the valuation of American options. However, Bauer et al. (2010) were the first to implement it in an SCR context, to the best of our knowledge. The main purpose of the Least-Squares Monte Carlo algorithm is to reduce the number of inner simulations, possibly even to one path. In the first phase, a regression function is constructed using these inner estimates. The accuracy of the inner estimates is drastically reduced by reducing the number of inner simulations, but, by combining the results of all outer simulations, the inner errors cancel out. In the second phase of the algorithm, this regression function is used to evaluate the conditional expectation at $t = 1$, without the need for inner simulations. For more details regarding the Least-Squares Monte Carlo algorithm, we refer the reader to Bauer et al. (2010).

Appendix B. Dynamic Heston Model

Based on the methodologies of [Alexander et al. \(2009\)](#), we assume \bar{v} and γ in Equation (13) to be stochastic in the hedge test, i.e.,

$$\begin{cases} d\bar{v}_t = \kappa_{\bar{v}}(\bar{v}_{\text{Mean}} - \bar{v}_t)dt + a_{\bar{v}}\bar{v}_t \left(\rho_{\bar{v}}dW_t^2 + \sqrt{1 - \rho_{\bar{v}}^2}dW_t^{\bar{v}} \right), \\ d\gamma_t = \kappa_{\gamma}(\gamma_{\text{Mean}} - \gamma_t)dt + a_{\gamma}\gamma_t \left(\rho_{\gamma}dW_t^2 + \sqrt{1 - \rho_{\gamma}^2}dW_t^{\gamma} \right), \end{cases} \tag{A1}$$

with the speed of mean reversion parameters $\kappa_{\bar{v}}$ and κ_{γ} , long-run averages \bar{v}_{Mean} and γ_{Mean} , volatilities $a_{\bar{v}}$ and a_{γ} and correlations $\rho_{\bar{v}}$ and ρ_{γ} . Moreover, $W_t^{\bar{v}}$ and W_t^{γ} are defined as independent Brownian motions. Note that, with the parameters $\rho_{\bar{v}}$ and ρ_{γ} close to 1, a high correlation with the volatility process is indicated, which is expected based on historical data. Under these assumptions, the option price is driven by the changes in \bar{v} and γ as well, giving

$$C_t^{\text{Dynamic}} \equiv C(t, S_t, v_t, r, \bar{v}_t, \gamma_t, \kappa, \rho, K, T). \tag{A2}$$

Next, we give details about the discretization of the dynamic Heston model. We can rewrite the dynamic Heston model with time-dependent \bar{v} and γ as follows

$$\begin{cases} dX_t = (r - \frac{1}{2}v_t)dt + \sqrt{v_t} \left(\rho dW_t^v + \sqrt{1 - \rho^2}dW_t^S \right), \\ dv_t = \kappa(\bar{v}_t - v_t)dt + \gamma_t \sqrt{v_t}dW_t^v, \\ d\bar{v}_t = \kappa_{\bar{v}}(\bar{v}_{\text{Mean}} - \bar{v}_t)dt + a_{\bar{v}}\bar{v}_t \left(\rho_{\bar{v}}dW_t^{\bar{v}} + \sqrt{1 - \rho_{\bar{v}}^2}dW_t^{\bar{v}} \right), \\ d\gamma_t = \kappa_{\gamma}(\gamma_{\text{Mean}} - \gamma_t)dt + a_{\gamma}\gamma_t \left(\rho_{\gamma}dW_t^{\gamma} + \sqrt{1 - \rho_{\gamma}^2}dW_t^{\gamma} \right), \end{cases} \tag{A3}$$

with $X_t = \log(S_t)$. We can simulate v_t and S_t with the Quadratic Exponential scheme proposed in [Andersen \(2008\)](#), with as minor difference that \bar{v} and γ are different in each time-step. The next step is to simulate \bar{v}_t and γ_t , such that they are correlated to v_t . First, we discretize the processes

$$\begin{cases} v_{t+\Delta t} \approx v_t + \kappa \left(\bar{v}_t - \frac{v_t + v_{t+\Delta t}}{2} \right) \Delta t + \gamma_t \sqrt{\frac{v_t + v_{t+\Delta t}}{2}} \Delta t Z^v, \\ \bar{v}_{t+\Delta t} \approx \bar{v}_t + \kappa(\bar{v}_{\text{Mean}} - \bar{v}_t)\Delta t + a_{\bar{v}}\rho_{\bar{v}}\bar{v}_t \sqrt{\Delta t} Z^{\bar{v}} + a_{\bar{v}}\bar{v}_t \sqrt{\Delta t(1 - \rho_{\bar{v}}^2)} Z^{\bar{v}} \\ \quad + \frac{1}{2}a_{\bar{v}}^2\bar{v}_t ((Z^{\bar{v}})^2 - dt), \\ \gamma_{t+\Delta t} \approx \gamma_t + \kappa(\gamma_{\text{Mean}} - \gamma_t)\Delta t + a_{\gamma}\rho_{\gamma}\gamma_t \sqrt{\Delta t} Z^{\gamma} + a_{\gamma}\gamma_t \sqrt{\Delta t(1 - \rho_{\gamma}^2)} Z^{\gamma} \\ \quad + \frac{1}{2}a_{\gamma}^2\gamma_t ((Z^{\gamma})^2 - dt), \end{cases} \tag{A4}$$

where Z^v , $Z^{\bar{v}}$ and Z^{γ} are independent standard normal distributed random variables. Now, we are able to derive an approximation for Z^v , given $v_{t+\Delta t}$,

$$\sqrt{\Delta t} Z^v \approx \frac{1}{\gamma_t \sqrt{\frac{v_t + v_{t+\Delta t}}{2}}} \left(v_{t+\Delta t} - v_t - \kappa \left(\bar{v}_t - \frac{v_t + v_{t+\Delta t}}{2} \right) \Delta t \right). \tag{A5}$$

This approximation can be substituted into Equation (A4), which ensures the correlation between v_t , \bar{v}_t and γ_t .

Appendix C. VIX Heston: UK and Europe

Appendix C.1. Calibrated Parameters

Appendix C.1.1. Parameters obtained from UK data

$$\Omega_t^{\text{Heston}}(X) = \begin{cases} \kappa_t & = 1.0, \\ v_{0,t} & = (-0.0014 + 0.0096 \cdot \text{VFTSE}_t)^2, \\ \bar{v}_t & = (0.0590 + 0.0110 \cdot \text{VFTSE}_{\text{filter}_t})^2, \\ \gamma_t & = 0.2556 + 0.0206 \cdot \text{VFTSE}_t, \\ \rho_t & = -0.6858. \end{cases} \tag{A6}$$

Appendix C.1.2. Parameters obtained from Europe data

$$\Omega_t^{\text{Heston}}(X) = \begin{cases} \kappa_t & = 1.0, \\ v_{0,t} & = (0.0013 + 0.0094 \cdot \text{VIX}_t)^2, \\ \bar{v}_t & = (0.0518 + 0.0100 \cdot \text{VIX}_{\text{filter}_t})^2, \\ \gamma_t & = 0.0571 + 0.0252 \cdot \text{VIX}_t, \\ \rho_t & = -0.6471. \end{cases} \tag{A7}$$

Appendix C.2. Predicted Parameters

Appendix C.2.1. Parameters predicted for UK

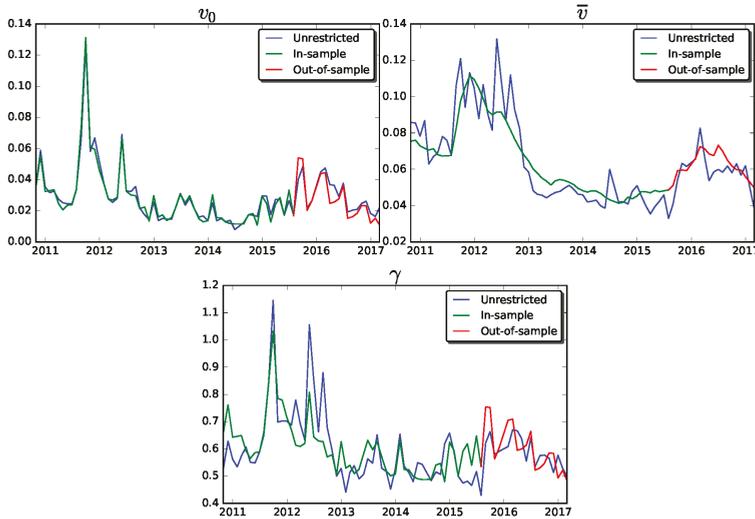


Figure A1. Prediction results Heston parameters of the UK dataset.

Appendix C.2.2. Parameters predicted for Europe

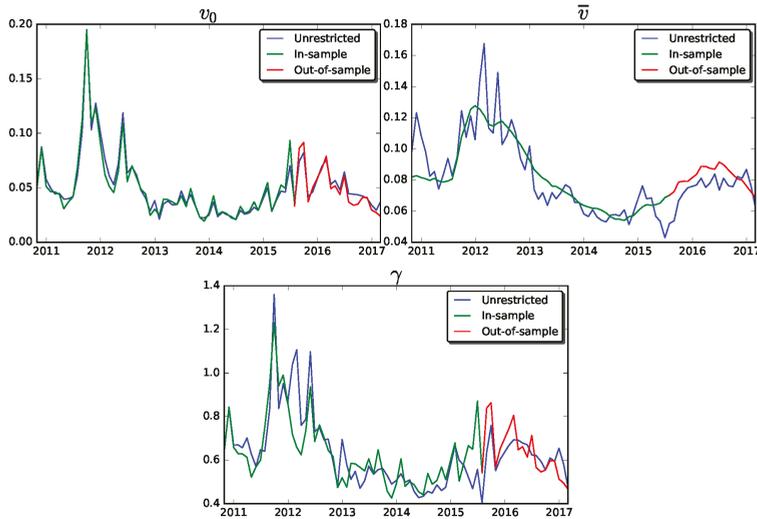


Figure A2. Prediction results Heston parameters of the Europe dataset.

References

Alexander, Carol, Andreas Kaeck, and Leonardo M. Nogueira. 2009. Model risk adjusted hedge ratios. *Journal of Futures Markets* 29: 1021–49. [CrossRef]

Andersen, Leif. 2008. Simple and efficient simulation of the Heston stochastic volatility model. *Journal of Computational Finance* 11: 1–42. [CrossRef]

Audrino, Francesco, and Dominik Colangelo. 2010. Semi-parametric forecasts of the implied volatility surface using regression trees. *Statistics and Computing* 20: 421–34. [CrossRef]

Bakshi, Gurdip, Charles Cao, and Zhiwu Chen. 1997. Empirical performance of alternative option pricing models. *The Journal of Finance* 52: 2003–49. [CrossRef]

Bauer, Daniel, Daniela Bergmann, and Andreas Reuss. 2010. Solvency II and nested simulations—A least-squares Monte Carlo approach. Paper presented at the 2010 ICA Congress, Cape Town, South Africa, March 7–12.

Bikker, Jacob A., and Haixia Hu. 2015. Cyclical patterns in profits, provisioning and lending of banks and procyclicality of the new Basel capital requirements. *PSL Quarterly Review* 55. Available online: <https://ojs.uniroma1.it/index.php/PSLQuarterlyReview/article/view/9907> (accessed on 10 October 2017).

CBOE. 2015. The CBOE Volatility Index—VIX: The Powerful and Flexible Trading and Risk Management Tool from the Chicago Board Options Exchange. Available online: <https://www.cboe.com/micro/vix/vixwhite.pdf> (accessed on 21 October 2018).

Cont, Rama, Jose da Fonseca, and Valdo Durrleman. 2002. Stochastic models of implied volatility surfaces. *Economic Notes* 31: 361–77. [CrossRef]

Devineau, Laurent, and Stéphane Loisel. 2009. Risk aggregation in Solvency II: How to converge the approaches of the internal models and those of the standard formula? *Bulletin Français d'Actuariat* 9: 107–45.

Duan, Jin-Chuan, and Chung-Ying Yeh. 2012. Price and Volatility Dynamics Implied by the VIX Term Structure Price and Volatility Dynamics Implied by the VIX Term Structure. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1788252 (accessed on 10 October 2017).

Gauthier, Pierre, and Pierre-Yves H. Rivaille. 2009. Fitting the Smile, Smart Parameters for SABR and Heston. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1496982 (accessed on 21 October 2017).

Hansen, Lars Peter. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50: 1029–54. [CrossRef]

Heston, Steven L. 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies* 6: 327–43. [CrossRef]

- Jain, Shashi, Patrik Karlsson, and Drona Kandhai. 2016. KVA, Mind Your P's and Q's! Available online: <https://ssrn.com/abstract=2792956> (accessed on 22 October 2017).
- Kenyon, Chris, Andrew David Green, and Mourad Berrahoui. 2015. Which Measure for PFE? The Risk Appetite Measure \mathbb{A}^* . Available online: <https://ssrn.com/abstract=2703965> (accessed on 21 November 2017).
- Longstaff, Francis A., and Eduardo S. Schwartz. 2001. Valuing American options by simulation: A simple least-squares approach. *The Review of Financial Studies* 14: 113–47. [CrossRef]
- Milevsky, Moshe A., and Thomas S. Salisbury. 2001. The real option to lapse a variable annuity: Can surrender charges complete the market. Paper presented at the 11th Annual International AFIR Colloquium, Toronto, ON, Canada, September 6–7.
- Mixon, Scott. 2002. Factors explaining movements in the implied volatility surface. *Journal of Futures Markets* 22: 915–37. [CrossRef]
- Pykhtin, Michael. 2012. Model foundations of the Basel III standardised CVA charge. *Risk* 25: 60.
- Ruiz, Ignacio. 2014. Backtesting counterparty risk: How good is your model? *Journal of Computational Finance* 10: 87–120. [CrossRef]
- Singor, Stefan N., Alex Boer, Josephine S. C. Alberts, and Cornelis W. Oosterlee. 2017. On the modelling of nested risk-neutral stochastic processes with applications in insurance. *Applied Mathematical Finance* 24: 1–35. [CrossRef]
- Stein, Harvey J. 2016. Fixing risk neutral risk measures. *International Journal of Theoretical and Applied Finance* 19: 1650021. [CrossRef]



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Journal of Risk and Financial Management Editorial Office
E-mail: jrfm@mdpi.com
www.mdpi.com/journal/jrfm



MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34
Fax: +41 61 302 89 18

www.mdpi.com



ISBN 978-3-03936-967-6